

УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“  
**ФАКУЛТЕТ ЗА ЕЛЕКТРОТЕХНИКА И ИНФОРМАЦИСКИ ТЕХНОЛОГИИ**  
ИНСТИТУТ ЗА КОМПЈУТЕРСКА ТЕХНИКА И ИНФОРМАТИКА

**Доне Стојанов**

**БИОИНФОРМАТИЧКА АНАЛИЗА НА ПОСТОЕЧКИ И  
НОВИ МОДЕЛИ НА ПРОТЕИНИ**

**- магистерски труд -**

**Ментор:**  
Академик проф.д-р Лупчо Коцарев

Скопје, Април 2010

# Содржина

Вовед.....	2
<b>1. Протеини.....</b>	<b>4</b>
1.1 Протеини-вовед.....	4
1.2 Структура на протеините.....	5
1.2.1 Примарна структура на протеините.....	9
1.2.2 Секундарна структура на протеините.....	10
1.2.3 Терциерна структура на протеините.....	11
1.2.4 Кавтернера структура на протеините.....	12
1.3 Генерална слика за структурата на протеините.....	12
1.4 Преглед на најупотребуваните модели и методи за просторно моделирање на структурата на протеините.....	14
1.4.1 Хомологно моделирање.....	16
1.4.2 АВ INITIO моделирање.....	17
1.5 Компјутерска визуелизација на протеински молекули.....	18
1.6 Споредување на протеини.....	20
1.7 Класификација на протеини.....	22
1.8 Предмет на истражување во рамки на магистерскиот труд.....	23
<b>2. Protein Data Bank.....</b>	<b>24</b>
<b>3. Просторно моделирање на протеините.....</b>	<b>28</b>
3.1 Предлог на просторен скелетен модел.....	28
3.2 Скелетен модел.....	29
3.3 Сублимат на карактеристиките на предложениот модел.....	40
<b>4. Просторно моделирање на конкретни протеини.....</b>	<b>41</b>
<b>5. Анализа на експериментално добиени податоци и примена на моделот.....</b>	<b>59</b>
5.1 Анализа на експериментално добиени податоци.....	59
5.2 Примена на предложениот скелетен модел.....	60
<b>6. Стохастички модел за конкретна структура на протеин од иста протеинска фамилија.....</b>	<b>64</b>
6.1 Предмет на анализа.....	64
6.2 Дискретни маркови вериги.....	65
6.2.1 Веројатности на n-чекорни премини и чапман-колмогорова равенка.....	66
6.2.2 Пристапливи состојби и иредуцибилни маркови вериги.....	67
6.2.3 Периодични и непериодични состојби на дискретна маркова верига.....	68
6.2.4 Рекурентни и транзиентни состојби на дискретна маркова верига.....	69
6.3 Опис на моделот.....	70
Заклучок.....	76

# Вовед

Протеините се едно од централните полиња на интерес, анализа и истражување во рамки на биоинформатиката. Таквиот пристап не е случаен и истиот е последица на улогата на истите во рамки на голем број на процеси кај живите организми. Сложената структура на протеините бара примена на точен, брз и ефикасен пристап на анализа, кој би бил невозможен без примена на математичко моделирање и компјутерски потпомогната обработка.

Појдовна точка во процесот на анализа на протеините е конструкцијата на модел кој точно и сеопфатно ќе ја опишува структурата на истите на одредено ниво на разгледување, при што таквиот модел мора да биде погоден за компјутерска репрезентација и обработка. По утврдувањето на структурата на протеинот на молекуларно ниво истата станува предмет на математичко моделирање, кое како краен резултат треба да даде математички дефиниран опис на физичката структура на протеинот на одредено ниво на разгледување. Таквиот опис во науката е дефиниран како математички модел, при што секој математички модел треба да биде точен, конзистентен и во целост математички да ја опишува физичката структура од интерес на бараното ниво на разгледување. Имајќи ја на ум сложената структура на протеините која имплицира сложени математички модели пропратени со огромен број на податоци и потребата од обработка на тие податоци со цел да се дојде до конкретни заклучоци во однос на истите, неминовно се наметнува потребата од компјутерска обработка и анализа. Процесот на компјутерска обработка и анализа на податоците започнува со изнаоѓање на погоден софтверски начин за претстава на податоците од математичките модели врз кои се применуваат ефикасни алгоритми за обработка кои резултираат со излезни податоци кои треба да бидат разбирливи за понатамошна научна анализа.

Во рамки на овој магистерски труд предложени се два нови пристапи на моделирање на структурата на протеините. Првиот предложен модел во рамки на овој магистерски труд ја опишува просторната структура на скелетот на протеинот и истиот е наречен скелетен модел. Моделот е изведен врз база на примена на концепти од векторска анализа и истиот претставува допринос кон низата на модели и методи за претстава на просторната структура на протеините. Моделирањето на просторната структура на протеините има голема важност во биоинформатиката, бидејќи структурата на протеините е тесно поврзана со нивната функционалност.

За разлика од скелетниот модел кој се базира на детерминистички пристап, вториот предложен модел се базира на стохастички пристап и истиот ја одредува веројатноста за конкретна примарна структура на протеин кој би припаѓал кон одредена протеинска фамилија. За таа цел се прави стохастичка маркова анализа на примарната структура на тековно откриените и верифицирани протеини од одредена протеинска фамилија, која на крај резултира со матрица на веројатности за промена на состојба од која се изведува генералната формула за веројатност на конкретна примарна структура на протеин кој би припаѓал кон таа протеинска фамилија.

Содржински магистерскиот труд е организиран во шест поглавја. Во рамки на првото поглавје протеините се разгледуваат како биомолекуларни ентитети со детален осврт на нивната градба и различни нивоа на структурирање, при што на крајот од поглавјето е даден преглед на најчесто употребените модели и методи за претстава на структурата на протеините. Тековно најголемата база на податоци во која се чуват детални податоци за сите откриени и верифицирани протеини Protein Data Bank е предмет на разгледување во рамки на второто поглавје. Во третото поглавје е дадена теоретската концепција на предложениот скелетен модел, кој веќе во четвртото поглавје е применет на конкретни протеини. Генералните резултати од примената на скелетниот модел на голем број протеини од различни протеински фамилии и заклучоците во однос на истите се дадени во петтото поглавје а во рамки на шесттото поглавје е образложен стохастичкиот модел за веројатност на конкретна примарна структура на протеин кој би припаѓал кон конкретна протеинска фамилија.

# 1

# Протеини

## 1.1 Протеини-вовед

Протеините се основни градбени единици и функционални елементи на живите организми. Протеините се инволвирани во сите биохемиски процеси кои се одвиваат во клетките. Како за пример протеините се градбени единици на ензимите, антителата, хормоните, транспортните молекули, косата, кожата, мускулите.

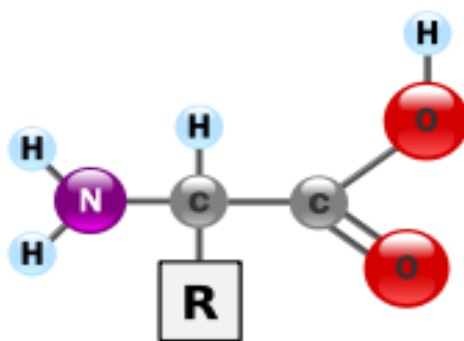
Протеините се молекули со сложена тродимензионална структура но истите истовремено можат да бидат разгледувани и како вериги од аминокиселини. Податоците за содржината на протеинот како линеарна верига од аминокиселини, функционалните аспекти на протеинот, организмот и делот од организмот кон кој тој специфичен протеин припаѓа како и информацијата за секундарната и просторната структура на протеинот се чуваат во протеинска база на податоци. Тековно најголемата база на податоци за чување на податоци за експериментално откриените протеински структури е Protein Data Bank(PDB), во која се чуваат податоци за примарната, секундарната и терциерната структура на протеините. Врз база на функционалните аспекти на протеинот или делот од организмот или клетката каде протеинот се појавува како градбен елемент протеините се класифицирани на протеински фамилии.

## 1.2 Структура на протеините

Протеините во основа се вериги од аминокиселини. На Слика 1.2.1 е прикажана структурата на една аминокиселина во општост. Сите аминокиселини имаат слична молекуларна структура која овозможува поврзување на аминокиселините во верига. Секоја аминокиселина задолжително содржи централен  $\alpha$  јаглероден атом  $C_\alpha$ , амино група  $NH_2$ , карбоксилна група  $COOH$  и еден водороден атом  $H$ . Аминокиселините меѓу себе се разликуваат по градбата на резидуумот  $R$  чија структура во основа ја определува самата аминокиселина.

Во градбата на протеините врз база на хемиската различност на структурата на резидуумот  $R$  влегуваат дваесет различни аминокиселини. Секоја аминокиселина се разликува од останатите со својата структура, големина како и електричните и хемиски својства кои ги поседува. Меѓу амино групата  $NH_2$  од една аминокиселина и карбоксилната група  $COOH$  од друга аминокиселина може да се формира пептидна врска со која се овозможува концептот на поврзување на аминокиселините во вериги.

Веригите на аминокиселини можат да бидат пептиди кои се дефинирани како вериги со помалку од четириесет аминокиселини или вериги со поголем број на аминокиселини кои се дефинирани како полипептиди. Протеините се структури составени од една или повеќе вериги од аминокиселини.



Слика 1.2.1 Структура на аминокиселина

Амино-киселините се делат на неесенцијални и есенцијални аминокиселини. Неесенцијалните аминокиселини се генерираат по природен пат во рамки на организмот. Есенцијалните аминокиселини не можат да бидат генерирани по природен пат во рамки на организмот. Потребата на луѓето од есенцијалните аминокиселини се задоволува со конзумација на одредени прехранбени продукти кои се богати со овој вид на аминокиселини. Во продолжение се дадени имињата на сите дваесет есенцијални и неесенцијални аминокиселини кои влегуваат во градбата на протеините, нивните ознаки, симболи и резидуумски структури.

Амино-киселина	Ознака	Симбол	Структура на резидуум
Alanine	ALA	A	$\text{—CH}_3$
Arginine	ARG	R	$\text{—CH}_2\text{—CH}_2\text{—CH}_2\text{—N—C}\begin{matrix} \text{=NH}_2 \\ \text{NH}_2 \end{matrix}$
Asparagine	ASN	N	$\text{—CH}_2\text{—C}\begin{matrix} \text{=O} \\ \text{NH}_2 \end{matrix}$
Aspartic acid	ASP	D	$\text{—CH}_2\text{—C}\begin{matrix} \text{=O} \\ \text{OH} \end{matrix}$
Cysteine	CYS	C	$\text{—CH}_2\text{—SH}$
Glutamic acid	GLU	E	$\text{—CH}_2\text{—CH}_2\text{—C}\begin{matrix} \text{=O} \\ \text{OH} \end{matrix}$
Glutamine	GLN	Q	$\text{—CH}_2\text{—CH}_2\text{—C}\begin{matrix} \text{=O} \\ \text{NH}_2 \end{matrix}$
Glycine	GLY	G	$\text{—H}$
Proline	PRO	P	$\begin{matrix} \text{O} \\ \parallel \\ \text{—HOC—C—CH}_2 \\ \diagup \quad   \\ \text{HN} \quad \text{CH}_2\text{—CH}_2 \\ \diagdown \quad   \\ \text{CH}_2 \end{matrix}$
Serine	SER	S	$\text{—CH}_2\text{OH}$

Табела 1.2.1 Листа на неесенцијални аминокиселини

Амино-киселина	Ознака	Симбол	Структура на резидиум
Isoleucine	ILE	I	$\begin{array}{c} \text{H} \\   \\ -\text{C}-\text{CH}_2-\text{CH}_3 \\   \\ \text{CH}_3 \end{array}$
Histidine	HIS	H	$\begin{array}{c} \text{HC}=\text{N} \\    \quad    \\ -\text{C}-\text{C}=\text{CH} \\   \quad   \\ \text{H}_2 \quad \text{N} \\   \\ \text{H} \end{array}$
Leucine	LEU	L	$\begin{array}{c} \text{CH}_3 \\   \\ -\text{CH}_2-\text{CH}_3 \\   \\ \text{CH}_3 \end{array}$
Methionine	MET	M	$\begin{array}{c} -\text{C}-\text{C}-\text{S}-\text{CH}_3 \\   \quad   \\ \text{H}_2 \quad \text{H}_2 \end{array}$
Lysine	LYS	K	$-\text{C}-\text{C}-\text{C}-\text{C}-\text{NH}_3 \\   \quad   \quad   \quad   \\ \text{H}_2 \quad \text{H}_2 \quad \text{H}_2 \quad \text{H}_2$
Phenylalanine	PHE	F	$-\text{C}-\text{C}_6\text{H}_5 \\   \\ \text{H}_2$
Tryptophan	TRP	W	$\begin{array}{c} -\text{C}-\text{C}=\text{CH} \\   \quad    \\ \text{H}_2 \quad \text{NH} \\   \\ \text{C}_6\text{H}_5 \end{array}$
Tyrosine	TYR	Y	$-\text{C}-\text{C}_6\text{H}_4-\text{OH} \\   \\ \text{H}_2$
Threonine	THR	T	$\begin{array}{c} \text{H} \\   \\ -\text{C}-\text{CH}_3 \\   \\ \text{OH} \end{array}$
Valine	Val	V	$\begin{array}{c} \text{CH}_3 \\ / \\ -\text{CH} \\ \backslash \\ \text{CH}_3 \end{array}$

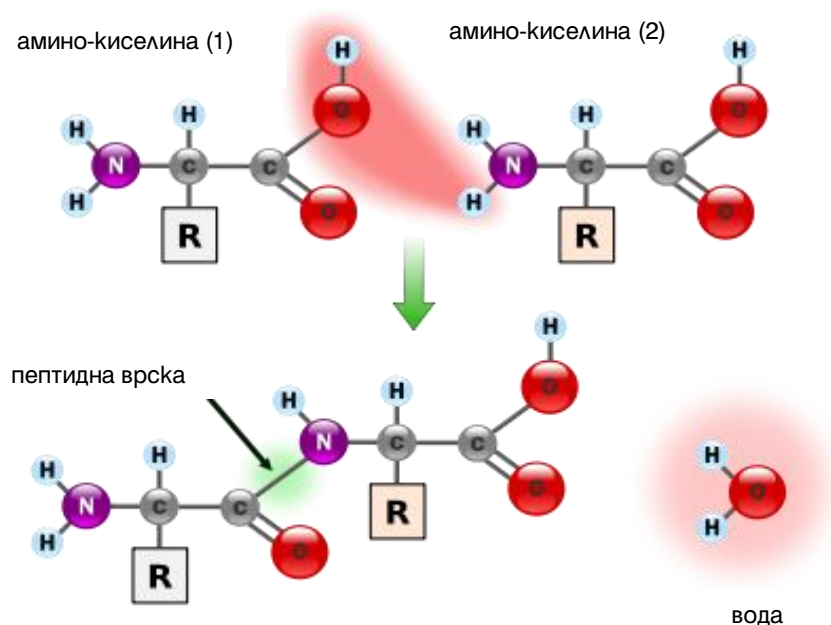
Табела 1.2.2 Листа на есенцијални аминокиселини



Амино-киселините учествуваат во одржувањето на стабилноста на протеините, стимулацијата на невротрансмитерите, регулацијата на шеќер во крвта, процесите на создавање и раст на хормоните, создавањето на антитела, формирањето на колаген, регулацијата на метаболизмот и се главни енергетски снабдувачи на мускулите по исцрпни физички активности.

Според структурата на резидуумот аминокиселините можат да бидат: неутрални(глицин, аланин, серин, треонин, метионин, цистеин, валин, леуцин, изолеуцин, пролин, фенилаланин, тирозин, триптофан), кисели(аспарагинска и глутаминска киселина) и базни(лизин, аргинин, хистидин, аспарагин, глутамин).

Секои две аминокиселини можат да формираат поголема молекула наречена дипептида со формирање на пептидна врска меѓу нив. Пептидната врска се формира меѓу јаглероден атом од карбоксилната **COOH** група и азотен атом од амино групата **NH<sub>2</sub>**. На Слика 1.2.2 е прикажано поврзувањето на две аминокиселини со ковалентна пептидна врска при што се ослободува и една молекула на вода **H<sub>2</sub>O**. Формираната пептидна врска меѓу било кои две аминокиселини може да се раскине само со дејство на соодветни ензими како што е ензимот трипсин. Низа од аминокиселини поврзани меѓу себе со пептидни врски претставува аминокиселинска верига а протеините се структури составени од една или повеќе аминокиселински вериги.



Слика 1.2.2 Поврзување на аминокиселини

Структурата на протеините може да биде разгледувана на примарно, секундарно, терциерно и кватернерно ниво и во продолжение ќе бидат дадени карактеристиките на организација на протеините на секое од овие нивоа.

### 1.2.1 Примарна структура на протеините

Примарната структура на протеините претставува концепт на разгледување на протеините како низа од аминокиселини, каде секоја позиција на таа низа -  $i$  е пополнета со некоја аминокиселина без разгледување на тродимензионалната структура на протеинот како и неговата просторна ориентација.

LYS	VAL	PHE	ARG	CYS	GLU	LEU	ALA
-----	-----	-----	-----	-----	-----	-----	-----

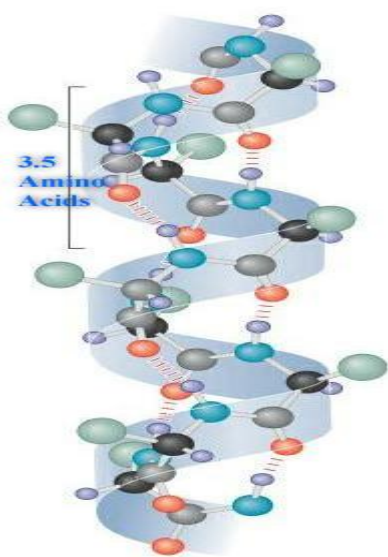
Слика 1.2.1.1 Примарна структура на протеините

## 1.2.2 Секундарна структура на протеините

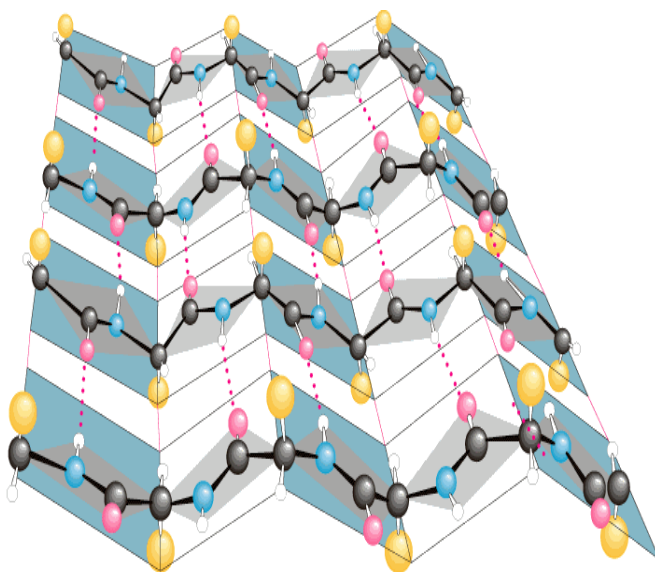
Ако се разгледа просторната структура на повеќе различни протеини може да се увиди дека постојат локални делови кај протеините кои се идентично геометриски структурирани. Тие идентични структурирања кај протеините се дефинирани со секундарната структура на протеините и истите се резултат на формирање на водородни врски меѓу кислороден атом од една аминокиселина и азотен атом од друга аминокиселина. Денес најстабилни два облици на секундарната структура на протеините се алфа хеликсите и бета рамнините.

Алфа хеликсот е стабилна структура која се воспоставува со формирање на водородни врски меѓу кислородните атоми од **CO** групите од аминокиселините на позиција  $k$  и азотните атоми од **NH** групите од аминокиселините на позиција  $k + 4$ . Таа е извиткана периодична структура која е прикажана на Слика 1.2.2.1, при што статистички е утврдено дека просечниот број на аминокиселини по периода е 3.5.

Структурата составена од две или повеќе паралелни поплипептидни вериги од аминокиселини која е формирана и стабилизирана со воспоставување на водородни врски меѓу кислородните и азотните атоми од секои две паралелни соседни аминокиселини е позната како бета рамнина. На Слика 1.2.2.2 е прикажана структурата на една бета рамнина.



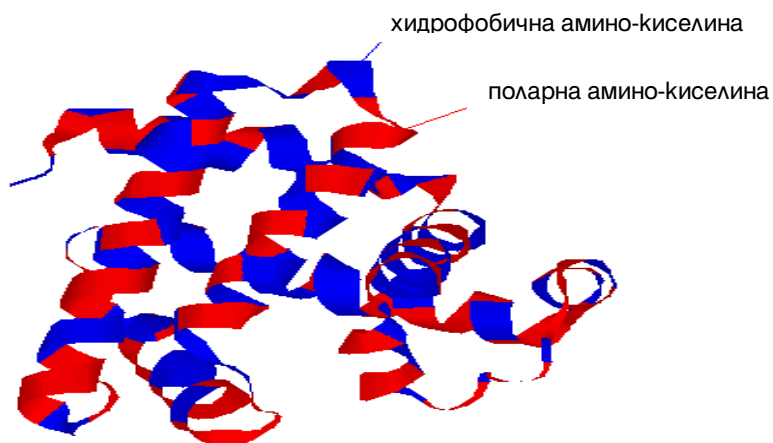
Слика 1.2.2.1 Алфа хеликс



Слика 1.2.2.2 Бета рамнина

### 1.2.3 Терциерна структура на протеините

Терциерната структура на протеините е резултат на различноста на резидиумите во рамки на аминокиселините. Дел од резидиумите кои влегуваат во рамки на дваесетте аминокиселини кои се градбени единици на протеините се поларни и хидрофилични(можат да стапат во реакција со вода) а пак останатите се неполарни и хидрофобични(не можат да стапат во реакција со вода). Меѓу спротивните резидиуми кај аминокиселините кои веќе се дел од некоја секундарна структура доаѓа до формирање на електронски врски кои ја дават крајната просторна структура на протеинот како тродимензионална структура составена од секундарни протеински компоненти(алфа хеликси, бета рамнини,...). Такви врски се: водородните врски кои можат да се воспостават меѓу два поларни резидиуми, врските кои се воспоставуваат меѓу два спротивно наелектризирани атоми од два резидиуми или дисулфидна врска која се смета за една од најсилните меѓурезидиумски врски. Терциерната структура на протеините ја одредува нивната функционалност и докажано е дека протеини со слична терциерна структура имаат слични функционалности. На Слика 1.2.3.1 е прикажана терциерната структура на еден протеин.



Слика 1.2.3.1 Терциерна структура на протеин

## 1.2.4 Кватернерна структура на протеините

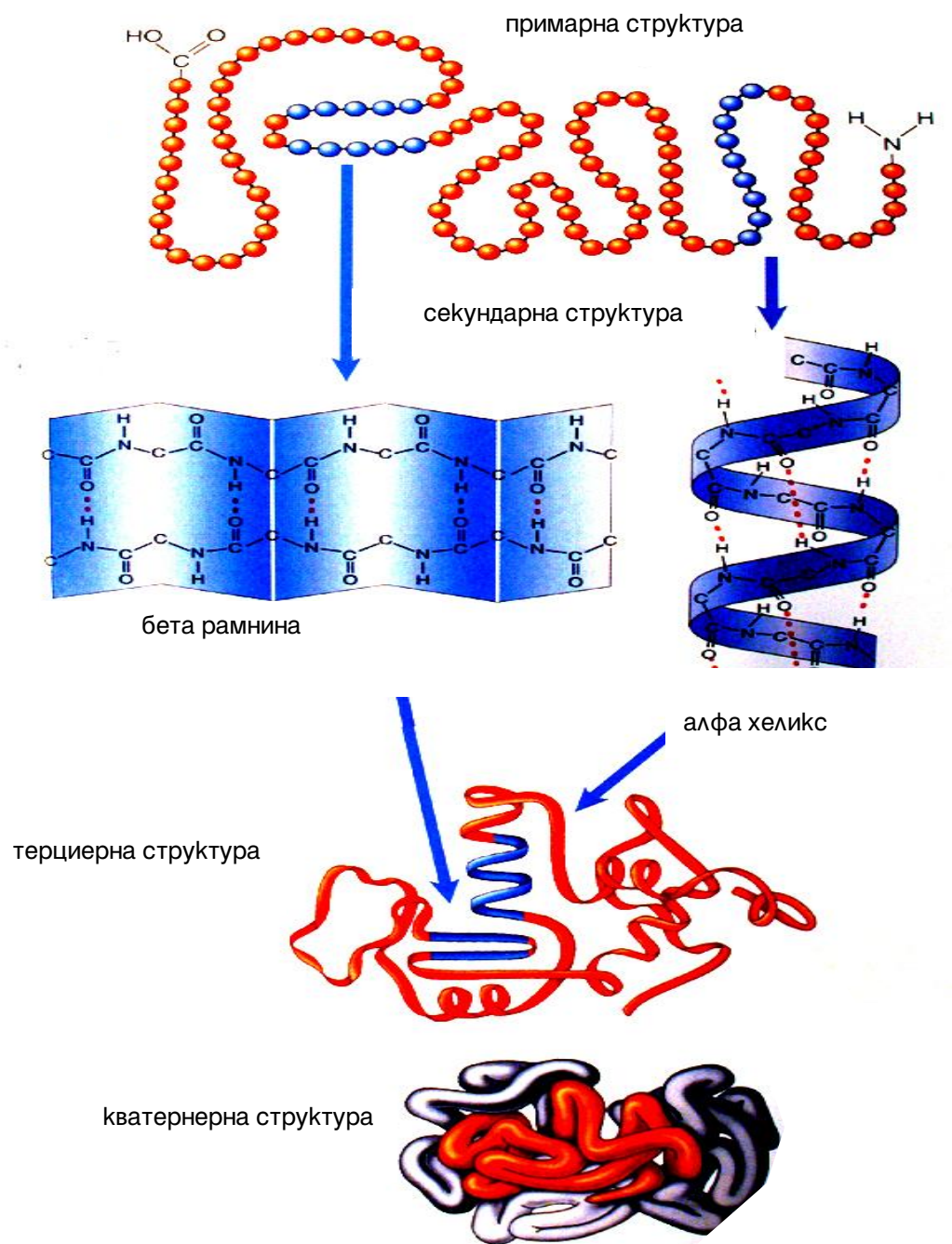
Голем дел од протеините се изградени од две или повеќе полипептидни вериги од аминокиселини наречени мономери. Најголем дел од протеините кои во свои рамки вклучуваат повеќе од 50 000 аминокиселини се изградени од два или повеќе нековалентно поврзани мономери. Поставеноста на овие мономери ја одредува кватернерната структура на протеините. Така на пример на Слика 1.2.4.1 е дадена кватернерната структура на протеинот хемоглобин која е одредена со поставеноста на четирите мономери од кои е изграден овој протеин, два  $\alpha$  мономери со по 141 аминокиселини и два  $\beta$  мономери со по 146 аминокиселини.



Слика 1.2.4.1 Кватернерна структура на протеинот хемоглобин

## 1.3 Генерална слика за структурата на протеините

Унијата од сите четири нивоа на разгледување на структурата на протеините најдобро ја опишува Слика 1.3.1. На истата е прикажана прво примарната структура на протеините како низа од аминокиселини групирани во пептидни ланци. Потоа поради формирање на водородни врски меѓу азотните и кислородните атоми од аминокиселините доаѓа до формирање на генерални облици на локално полипептидно извиткување како што се алфа хеликсите и бета рамнините, кои ја определуваат секундарната структура на протеините. Наредно поради различната поларизираност на резидуумите доаѓа до формирање на електронски врски меѓу спротивните резидууми кои ја оформуваат конечната просторна терциерна структура на протеините, додека пак множеството на нековалентно поврзани мономери ја определува кватернерната структура на протеините.



Слика 1.3.1 Унија од сите нивоа на разгледување на структурата на протеините

## 1.4 Преглед на најупотребуваните модели и методи за просторно моделирање на структурата на протеините

Во основа постојат два генерални методи за моделирање на просторната структура на протеините. Тоа се компаративното моделирање и *de novo* пристапите. Компаративното моделирање ја предвидува просторната структура на протеинот со споредување на неговата примарна структура со примарните структури на протеини со познати просторни структури. Компаративното моделирање се базира на поставката дека слична примарна структура на два протеини резултира со слична просторна структура на истите. Сличност на примарната структура од 50% или повеќе на споредуваниот протеин со референтниот протеин најчесто резултира со предвиден просторен модел за споредуваниот протеин со голема точност. Најчесто употребуван алгоритам за наоѓање на протеински структури со слична примарна структура е BLAST алгоритмот, при што ако резултатот од пребарувањето се повеќе протеини со слични примарни структури, тогаш се избира протеинот со најголем процент на примарна структурна сличност со споредуваниот чија просторна структура е позната.

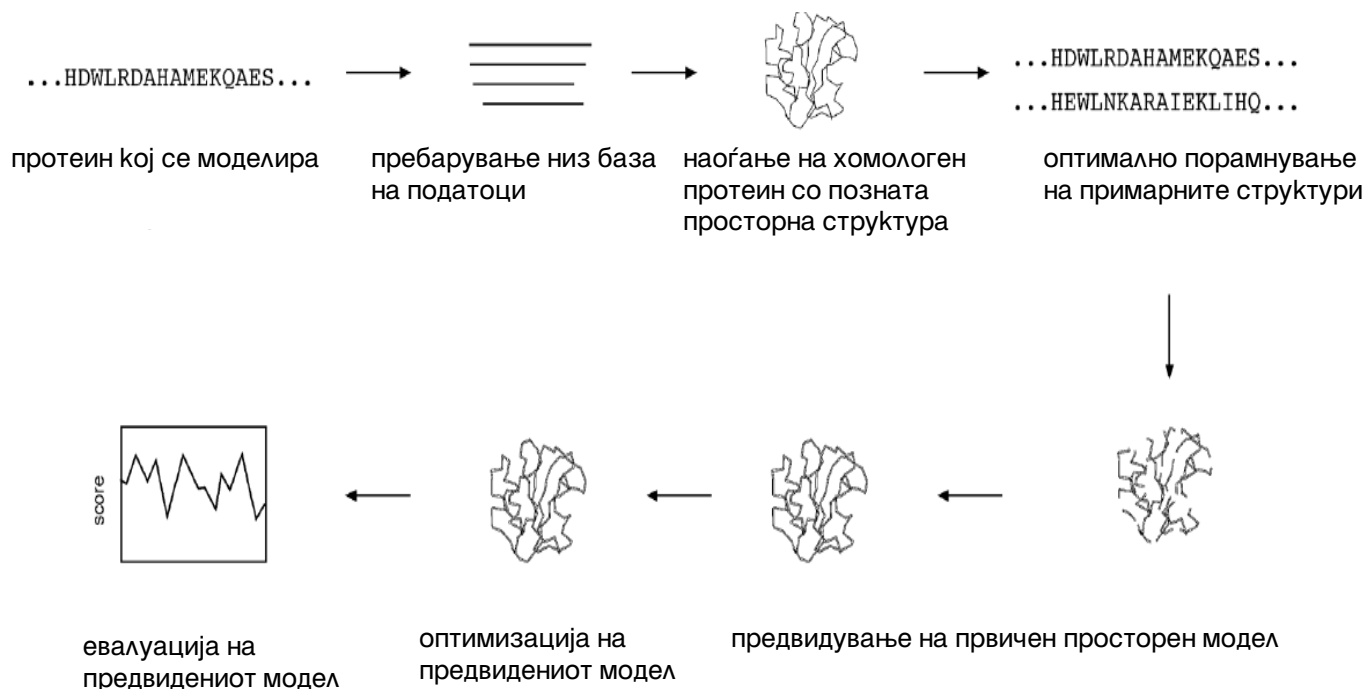
*De novo* пристапите на моделирање на протеините за разлика од компаративното моделирање немаат потреба од постоење на споредбени протеини со познати просторни структури, и тоа ги прави применливи за секој протеин. Некои *de novo* методи ја предвидуваат просторната структура на протеинот земајќи ги во предвид меѓуенергетските аспекти во рамки на протеинот. Такви се *ab initio* методите кои ја предвидуваат просторната структура на протеините врз база на утврдување на низа на енергетски функции кои ги опишуваат физичко-хемиските интеракции меѓу атомите во рамки на протеинот. *Ab initio* методите се пропратени со голем број на комплексни математички пресметки и до денес истите се покажаа како успешни во предвидувањето на просторната структура на протеини со мал број на аминокиселини и во предвидувањето на просторната структура на мали сегменти од подолги протеински секвенци. Најуспешните *de novo* методи кои се применливи на подолги протеински секвенци (на протеини со должина до 150 аминокиселини) ја предвидуваат просторната структура на протеинот врз база на веќе утврдени просторни протеински структурирања.

Во последната декада е направен огромен напредок на полето на откривање на нови de novo методи кои ќе дадат точни резултати. Изненадувачка е огромната точност на некои од поновите de novo методи во предвидувањето на просторната структура на помали протеини и на локални сегменти од подолги протеински фамилии, но сеуште нивната примена на протеини со поголем број на аминокиселини е непрактична. Ако се направи споредба во однос на точноста на компаративните методи и de novo методите може да се утврди дека компаративните методи тековно се во предност, што ги прави најупотребувани во биоинформатичката протеинска структурна анализа.



## 1.4.1 Хомологно моделирање

Хомологното моделирање е процес кој се одвива во повеќе фази. Прво се врши пребарување низ протеинска база на податоци за да се најдат хомологни протеини со познати просторни структури. Во овој случај за хомологни протеини се земаат протеини со слични примарни структури со минимум 30% верижна идентичност. Резултатот од пребарувањето кое е имплементирано софтверски со алгоритми како BLAST или FASTA, може да даде повеќе хомологни протеини, при што во тој случај се одбира протеинот со најголема примарна структурна сличност со протеинот кој просторно се моделира. По наоѓањето на хомологен протеин со најголем процент на примарна структурна сличност се применува алгоритам за оптимално порамнување на примарните структури на протеинот кој се моделира и хомологниот протеин. Со примена на оптимално порамнување се одредува максимум можни еквивалентни резидиумски парови меѓу протеинот предмет на моделирање и хомологниот протеин, по што следи копирање на координатните скелетни и еквивалентни резидиумски позиции од хомологниот протеин во слободниот простор резервиран за приказ на просторната структура на протеинот предмет на моделирање. Наредно се одредуваат координатните позиции на резидиумите кај протеинот предмет на моделирање кои не формираат еквивалентни резидиумски парови со резидиумите од хомологниот протеин, и истите се одредуваат со анализа на соседните меѓурезидиумски взаемни дејствија. Реалната просторна структура на протеинот е резултат на минимален енергетски меѓуатомски баланс, при што ако тој концепт се примени на тековно предвидениот просторен модел за протеинот кој просторно се моделира можно е да се коорегираат одредени неправилности за предвидените просторни протеински координатни позиции со што се прави оптимизација на првично предвидениот просторен модел. Во завршната фаза предвидениот оптимален структурен модел се евалуира со проверка на одредени параметри како што се протеинските меѓуатомски растојанија, просторни агли,...итн.



Слика 1.4.1.1 Приказ на процесот на хомологно моделирање

## 1.4.2 АВ INITIO моделирање

За да се предвиди просторната структура на протеин со позната примарна структура со хомологно моделирање неопходно е да постои барем еден хомологен протеин(протеин со слична примарна структура) со позната просторна структура за протеинот кој просторно се моделира. Ако пребарувањето на хомологни протеини не даде ниту еден позитивен резултат хомологното моделирање е целосно неприменливо. АВ INITIO моделирањето нема потреба од постоење на протеини со познати просторни структури и истото ја предвидува просторната структура на протеинот врз база на утврдени просторни структурирања на делови од протеини. Типични АВ INITIO методи за предвидување на просторната структура на протеини се Chou-Fasman алгоритмот и GOR методот.

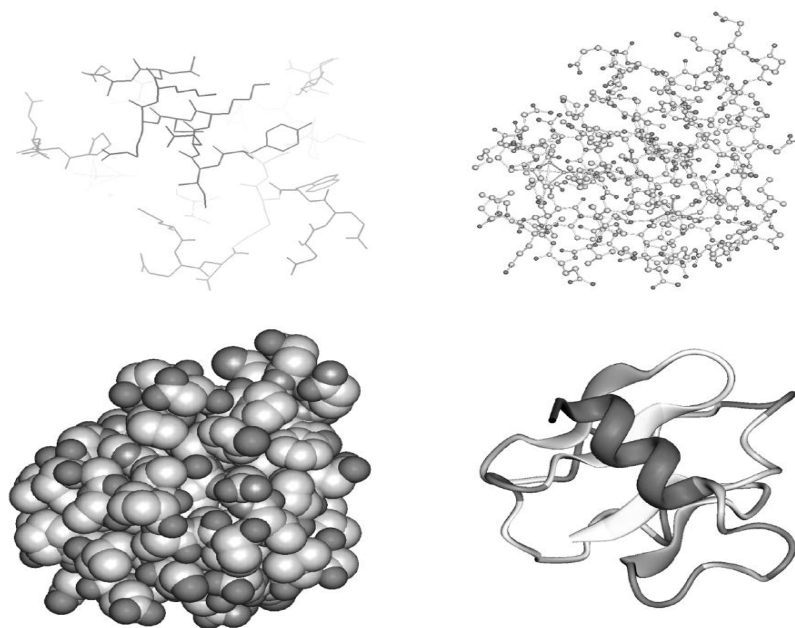
Chou-Fasman алгоритмот врши предвидување на просторната структура на протеинот со утврдување на видовите на секундарни структурирања во рамки на кои најчесто се среќаваат двесетте градебни аминокиселини. Така утврдено е дека аланинот, глутаминската киселина и метионинот најчесто се дел од алфа хеликс, додека пак не е многу веројатно аминокиселините глицин и пролин да бидат составен дел на алфа хеликс.

GOR методот ја предвидува просторната структура на протеин со утврдување на припадност на протеинските подсеквенци со должина од 17 аминокиселини кон едно од четирите секундарни структурирања: хеликс(H), стандард(E), извиткување(T) или навивка(C).

## 1.5 Компјутерска визуелизација на протеински молекули

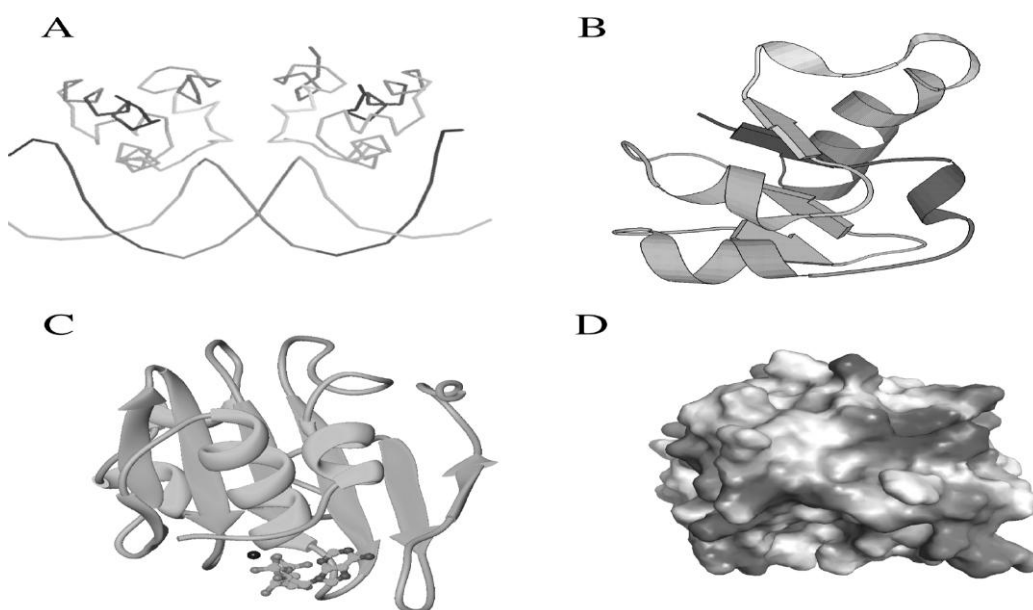
Секоја компјутерска програма за визуелизација на структурата на протеините имплементира еден или повеќе различни молекуларни визуелизациски формати. Молекуларниот визуелизациски формат воспоставува релација меѓу протеинските атоми, и истиот овозможува приказ на протеинската молекула од одреден аспект на разгледување. Псофистицираните програми за визуелизација на структурата на протеините имплементираат повеќе молекуларни визуелизациски формати со што се овозможува приказ на структурата на протеинот од различни аспекти. Едни од најчесто користените молекуларни визуелизациски формати за приказ на структурата на протеините се следниве:

- Молекуларен визуелизациски формат кој се задржува само на приказ на врските меѓу атомите кои се претставуваат со отсечки.
- Молекуларен визуелизациски формат кој дава приказ на протеинската молекула како множество на атоми со меѓусебни врски, каде атомите се претставени со сфери а меѓуатомските врски со отсечки.
- Молекуларен визуелизациски формат кој дава приказ на протеинската молекула како множество на атоми со ван дер Валсови радиуси претставени со сфери со пропорционална распределба на радиусите на сферите во претставата.
- Молекуларен визуелизациски формат кој дава приказ на просторната распределба на секундарните структурирања кај протеинската молекула.



**Слика 1.5.1** Приказ на претходните четири молекуларни визуелизациски формати

Најчесто користени софтверски апликации за приказ на молекуларната структура на протеините се: RasMol, Swiss-PDBViewer, Molscript, Ribbons, Grasp,... итн. Сите претходни софтверски апликации како влез ги примаат координатите на протеинските атоми по што следи воспоставување на релација меѓу истите, примена на еден или повеќе молекуларни визуелизациски формати кои резултираат со една или повеќе различни претстави за молекуларната структура на протеинот. На Слика 1.5.2 е даден молекуларниот приказ на протеини кај различни софтверски апликации.



**Слика 1.5.2** Молекуларен приказ на протеини кај: (A) Rasmol, (B) Molscript, (C) Ribbons, (D) Grasp

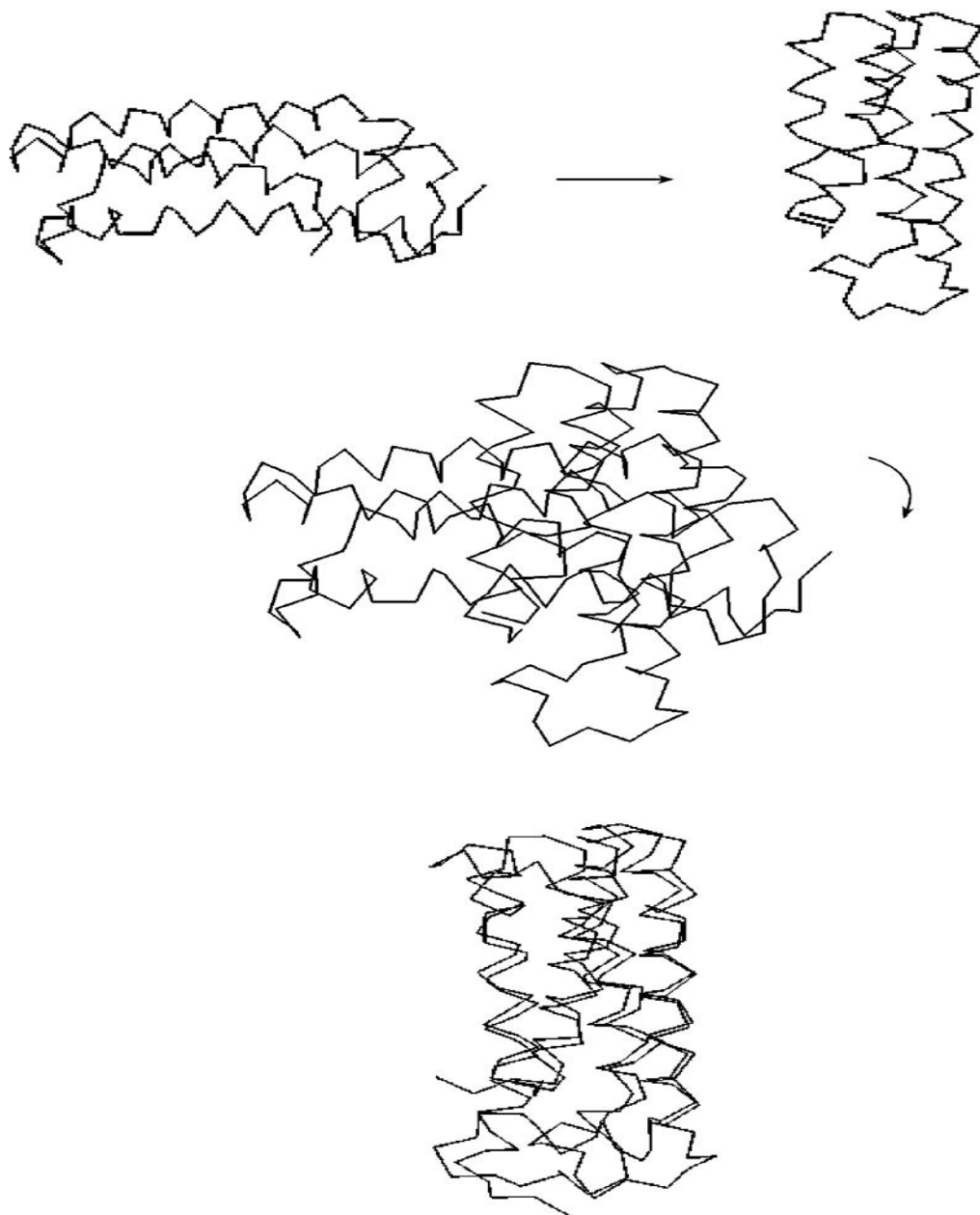
## 1.6 Споредување на протеини

Споредувањето на два протеини е процес насочен кон утврдување на сличности помеѓу истите. Процесот на споредување на протеините најчесто вклучува директно порамнување на нивната примарна структура и просторна споредбена анализа со која се утврдуваат деловите кај протеините кои се идентично просторно структурирани и деловите кои се различно просторно структурирани. Два протеини може да имат слична просторна структура и без сличност на нивните примарни структури. Таквата просторна сличност укажува на подалечна хомологија (евулуциона релација) помеѓу тие протеини, и истата не може да се утврди само со примена на порамнување на нивните примарни структури. Споредувањето на просторната структура на протеините најчесто се прави алгоритамски со изнаоѓање на максимална просторна сличност помеѓу споредуваните протеини, со примена на интермолекуларниот или интрамолекуларниот метод, а понекогаш се употребува и хибриден метод од претходните два.

Интермолекуларниот метод најчесто се применува за споредување на протеини со слична структура. Процесот започнува со утврдување на еквивалентни резидиумски парови по што следи трансформација на еден од протеините кон другиот така што двата протеини се доведуваат во иста координатна рамка. Откако протеините ќе се доведат во иста координатна рамка поместениот протеин се ротира во однос на референтниот протеин, при што постојано се мерат растојанијата помеѓу централните јаглеродни атоми  $C_{\alpha}$  кај еквивалентните резидиумски парови. Процесот на ротација завршува кога ќе се постигне минимално меѓумолекуларно растојание кое одговара на минимум на

функцијата  $f = \sqrt{\frac{\sum_{i=1}^n D_i^2}{N}}$ , каде  $N$  го означува бројот на еквивалентни резидиумски парови а со  $D_i$  со

означени растојанијата помеѓу централните јаглеродни атоми кај паровите еквивалентни резидиуми. Меѓусебната просторна положба на протеинските молекули за која меѓумолекуларното растојание е минимално според овој метод одговара на максимално просторно совпаѓање на протеинските молекули.



**Слика 1.6.1** Приказ на интермолекуларниот метод за утврдување на максимално просторно совпаѓање помеѓу два протеини

За разлика од интермолекуларниот метод кој се применува за утврдување на максимално просторно совпаѓање на два протеини со слична примарна структура, интрамолекуларниот метод генерално се применува за протеини со различна примарна структура. Појдовно кај интрамолекуларниот метод за секој протеин кој просторно се споредува се составува матрица на резидиумски растојанија. Матриците на резидиумски растојанија потоа се споредуваат за да се утврди максимално меѓусебно совпаѓање. Со примена на овој метод се утврдуваат регионите кај протеините кои се слично просторно структурирани.

## 1.7 Класификација на протеини

Резултатот од споредувањето на протеините се меѓусебно класифицирани протеини. Класификацијата на протеини воспоставува хиерархиска релација помеѓу протеините и дава еволуциски поглед на развојот на протеинските структури. Класификацијата на протеини во денешно време се имплементира софтверски, при што најпознати два системи за класификација на протеини се системите SCOP и CATH.

Протеините кај SCOP (Structural Classification Of Proteins) системот се организирани хиерархиски во класи, извиткувачки класи, суперфамилии и фамилии. Фамилиите кај SCOP системот се состојат од протеини со слична примарна структура, која изразена во бројки изнесува повеќе од 30% верижна идентичност. Суперфамилиите се состојат од фамилии со голема структурна сличност, но со мала сличност на нивните примарни структури. Се верува дека членовите на суперфамилиите имаат далечен заеднички претходник и истите функционално се разликуваат. Множество на суперфамилии со слични секундарни структурирања пропратени со слични просторни ориентации и врски формира извиткувачка класа, при што не секогаш членовите на извиткувачката класа имаат заедничка еволуциона релација. Највисокото ниво во хиерархијата на поделба на видовите на класификации на протеини кај SCOP системот го завзема класификацијата на протеините исклучиво според нивната секундарна структура. Таквата класификација ги групира протеините во класи, при што постои класа на протеини составени само од алфа хеликси, класа на протеини составени само од бета рамнини, класа на протеини составени од алфа хеликси и бета рамнини,...итн.

Кај CATH (Class, Architecture, Topology and Homologous) системот протеините се организирани во класи, архитектури, топологии, хомологни суперфамилии и хомологни фамилии. Хомологните фамилии и хомологните суперфамилии идентично се дефинирани како кај SCOP системот. Топологија кај CATH е синоним за тоа што беше дефинирано како извиткувачка класа кај SCOP системот. Архитектура е класификациско ниво карактеристично само за CATH и истото ја опишува распределбата на секундарните структури независно од нивната меѓусебна поврзаност. Класификациското ниво класа кај CATH системот е идентично дефинирано како кај SCOP системот.

## 1.8 Предмет на истражување во рамки на магистерскиот труд

Протеините се системи од аминокиселински атоми. Меѓу атомите кај протеините постојат низа на врски, сили и взаемнодејствија кои ја определуваат просторната структура на протеините. Бројот на атоми во рамки на протеините е многу голем а следствено на тоа и бројот на атомски взаемни дејствија исто така ќе биде многу голем. Анализата на секое од тие взаемни дејствија во практични услови речиси е невозможна и затоа протеините кога се разгледуваат како системи од атоми се разгледуваат како концепти од теоријата на хаосот. Атомите кои учествуваат во градбата на протеините се дадени во Табела 1.8.1 .

Атом	Симбол
јаглероден атом	C
азотен атом	N
кислороден атом	O
водороден атом	H
сулфурен атом	S

Табела 1.8.1 Протеински атоми

Предмет на истражување и анализа на овој магистерски труд прво ќе биде да се предложи модел за просторната структура на скелетот на протеините, кој подоцна ќе биде применет на протеини од различни протеински фамилии. Откако ќе бидат утврдени генералните параметри на предложениот модел кој ја моделира структурата на скелетите на протеините на молекуларно ниво, ќе се предложи нареден стохастички модел за веројатност на конкретна примарна структура на протеин кој би припаѓал кон одредена протеинска фамилија, базиран на разгледување на протеините како маркови системи.



# Protein Data Bank

Податоците за структурата на протеините се чуваат во јавна база на податоци Protein Data Bank-PDB. Откако структурата на протеинот ќе се утврди главно со користење на ренгентска кристалографија или NMR спектроскопија, истата оди на валидација по што доколку исходот од валидацијата биде позитивен, податоците за новиот протеин се зачувуваат во pdb датотека.

Структурата на најголемиот дел од протеините за кои се чуваат податоци во Protein Data Bank е откриена со помош на ренгентска кристалографија а со NMR спектроскопија е откриена структурата на околу 15% од протеините за кои се чуваат податоци во Protein Data Bank.

Во Табела 2.1 е дадена детална статистика за откриените протеински структури во PDB базата на податоци до 10 Февруари 2010.

Метод	Тип на молекула				Вкупно
	протеини	нуклеински киселини	протеин / НА структури	останато	
Ренгентска кристалографија	51042	1191	2346	17	54596
NMR спектроскопија	7193	890	152	7	8242
Електронска микроскопија	184	17	70	0	271
Хибридизација	18	1	1	1	21
Друго	120	4	4	13	141
Вкупно	58557	2103	2573	38	63271

Табела 2.1 Статистика за откриени протеински структури до 10 Февруари 2010

Откако податоците за новооткриената протеинска структура ќе бидат зачувани во pdb датотека, истата се асоцира со секвенца од четири карактери која претставува идентификатор на истата. Така на пример податоците за човековиот протеин S100B во комплекс со S45 се зачувани во pdb датотека асоцирана со код за идентификација 3hcm и кон истата може да се пристапи преку URL: <http://www.rcsb.org/pdb/files/3hcm.pdb>.

Кај секоја pdb датотека можат да се издвојат следниве четири последователни фрагменти:

- Фрагмент во кој се чуваат податоци за името на структурата, бројот на ланци во рамки на структурата, функционални информации за структурата, изворот на структурата како и податоци за експерименталните техники кои се употребени за утврдување на структурата.

Во продолжение е прикажан таквиот фрагмент од pdb датотека за човековиот S100B протеин.

```
HEADER      METAL BINDING PROTEIN                                06-MAY-09   3HCM
TITLE       CRYSTAL STRUCTURE OF HUMAN S100B IN COMPLEX WITH S45
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: PROTEIN S100-B;
COMPND      3 CHAIN: A, B;
COMPND      4 SYNONYM: S100 CALCIUM-BINDING PROTEIN B, S-100 PROTEIN
COMPND      5 SUBUNIT BETA, S-100 PROTEIN BETA CHAIN;
COMPND      6 ENGINEERED: YES
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE      3 ORGANISM_COMMON: HUMAN;
SOURCE      4 ORGANISM_TAXID: 9606;
SOURCE      5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      6 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE      7 OTHER_DETAILS: THE PROTEIN WAS PURCHASED FROM A COMMERCIAL
SOURCE      8 SOURCE.
KEYWDS      S100B, CALCIUM BINDING PROTEIN, INHIBITOR, CALCIUM,
KEYWDS      2 CYTOPLASM, METAL-BINDING, NUCLEUS
EXPDTA      X-RAY DIFFRACTION
AUTHOR      S.MANGANI,L.CESARI
REVDAT      1 02-FEB-10 3HCM 0
```

---

- Фрагмент во кој се чуваат податоци за бројот на пептидни ланци во рамки на структурата и нивната содржина како вериги од аминокиселини.

Во продолжение е прикажан таквиот фрагмент од pdb датотека за човековиот S100B протеин.

```

SEQRES      1 A      92  MET SER GLU LEU GLU LYS ALA MET VAL ALA LEU ILE ASP
SEQRES      2 A      92  VAL PHE HIS GLN TYR SER GLY ARG GLU GLY ASP LYS HIS
SEQRES      3 A      92  LYS LEU LYS LYS SER GLU LEU LYS GLU LEU ILE ASN ASN
SEQRES      4 A      92  GLU LEU SER HIS PHE LEU GLU GLU ILE LYS GLU GLN GLU
SEQRES      5 A      92  VAL VAL ASP LYS VAL MET GLU THR LEU ASP ASN ASP GLY
SEQRES      6 A      92  ASP GLY GLU CYS ASP PHE GLN GLU PHE MET ALA PHE VAL
SEQRES      7 A      92  ALA MET VAL THR THR ALA CYS HIS GLU PHE PHE GLU HIS
SEQRES      8 A      92  GLU
SEQRES      1 B      92  MET SER GLU LEU GLU LYS ALA MET VAL ALA LEU ILE ASP
SEQRES      2 B      92  VAL PHE HIS GLN TYR SER GLY ARG GLU GLY ASP LYS HIS
SEQRES      3 B      92  LYS LEU LYS LYS SER GLU LEU LYS GLU LEU ILE ASN ASN
SEQRES      4 B      92  GLU LEU SER HIS PHE LEU GLU GLU ILE LYS GLU GLN GLU
SEQRES      5 B      92  VAL VAL ASP LYS VAL MET GLU THR LEU ASP ASN ASP GLY
SEQRES      6 B      92  ASP GLY GLU CYS ASP PHE GLN GLU PHE MET ALA PHE VAL
SEQRES      7 B      92  ALA MET VAL THR THR ALA CYS HIS GLU PHE PHE GLU HIS
SEQRES      8 B      92  GLU

```

---

- Фрагмент кој ја опишува секундарната структура на протеинот во кој се сместуваат податоци за видовите на секундарни структурирања и нивниот број во рамки на протеинот. Во овој фрагмент се чуват и податоци за тоа која е почетна а која завршна аминокиселина на секое секундарно структурирање.

Дел од тој фрагмент за човековиот S100B протеин е даден во продолжение.

```

HELIX       1      1 SER A      1  GLY A      19      1          19
HELIX       2      2 LYS A     28  LEU A     40      1          13
HELIX       3      3 GLU A     49  ASP A     61      1          13
HELIX       4      4 ASP A     69  GLU A     86      1          18
HELIX       5      5 SER B      1  GLY B     19      1          19
HELIX       6      6 LYS B     28  LEU B     40      1          13
HELIX       7      7 GLU B     49  ASP B     61      1          13
HELIX       8      8 PHE B     70  PHE B     88      1          19

```

---

- Фрагмент во кој се чуваат податоци за просторните координати на секој атом од протеинот.

Секој запис во тој дел започнува со зборот ATOM по што следуваат податоци за:

- редниот број на атомот во рамки на протеинот
- видот на атомот
- името на аминокиселината која го содржи атомот
- пептидниот ланец кон кој припаѓаат атомот и аминокиселината
- бројот на аминокиселината во рамки на структурата

Веднаш по овој дел се дадени просторните координати на атомот во облик  $(x, y, z)$  изразени во единици ангстреми  $1\text{\AA} = 10^{-10}\text{ m}$ , по што следат последните два податоци за радиусот на атомот и неговиот температурен фактор.

Во продолжение е прикажан дел од таквиот фрагмент од pdb датотека за човековиот S100B протеин.

ATOM	1	N	MET	A	0	16.174	-7.193	12.900	1.00	41.62	N
ATOM	2	CA	MET	A	0	14.846	-6.710	13.401	1.00	40.86	C
ATOM	3	C	MET	A	0	14.733	-5.190	13.305	1.00	39.59	C
ATOM	4	O	MET	A	0	15.691	-4.454	13.590	1.00	40.54	O
ATOM	5	CB	MET	A	0	14.606	-7.195	14.828	1.00	41.52	C
ATOM	6	CG	MET	A	0	14.203	-8.656	14.947	1.00	44.13	C
ATOM	7	SD	MET	A	0	12.421	-8.817	15.138	1.00	46.04	S
ATOM	8	CE	MET	A	0	12.133	-10.518	14.604	1.00	43.75	C
ATOM	9	N	SER	A	1	13.571	-4.707	12.883	1.00	37.78	N
ATOM	10	CA	SER	A	1	13.389	-3.258	12.700	1.00	36.06	C
ATOM	11	C	SER	A	1	13.094	-2.597	14.044	1.00	35.55	C
ATOM	12	O	SER	A	1	12.797	-3.284	15.022	1.00	35.10	O
ATOM	13	CB	SER	A	1	12.246	-2.969	11.718	1.00	36.56	C
ATOM	14	OG	SER	A	1	10.985	-3.224	12.323	1.00	31.21	O
ATOM	15	N	GLU	A	2	13.169	-1.274	14.078	1.00	34.39	N
ATOM	16	CA	GLU	A	2	12.928	-0.508	15.296	1.00	34.87	C
ATOM	17	C	GLU	A	2	11.516	-0.719	15.845	1.00	33.68	C
ATOM	18	O	GLU	A	2	11.338	-1.072	17.016	1.00	34.50	O
ATOM	19	CB	GLU	A	2	13.225	0.975	15.049	1.00	35.68	C
ATOM	20	CG	GLU	A	2	14.686	1.198	14.682	1.00	37.62	C
ATOM	21	CD	GLU	A	2	15.610	0.293	15.491	1.00	41.37	C
ATOM	22	OE1	GLU	A	2	15.666	0.451	16.740	1.00	42.46	O
ATOM	23	OE2	GLU	A	2	16.250	-0.595	14.882	1.00	43.85	O

# ПРОСТОРНО МОДЕЛИРАЊЕ НА ПРОТЕИНИТЕ

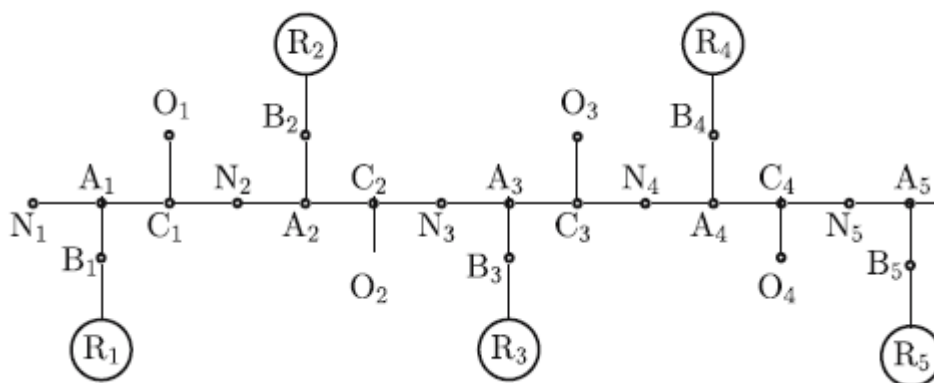
## 3.1 Предлог на просторен скелетен модел

Утврдено е дека просторната структура на протеинот ја одредува неговата функционалност. Со примена на фундаментална анализа во однос на протеините се доаѓа до заклучокот дека просторната структура на протеините на макроскопско ниво е резултат на голем број на хаотични атомски взаемни дејствија кои се одвиваат на микроскопско ниво. Просторните протеински модели предложени до сега главно се насочени кон репрезентација на просторната структура на протеините на макроскопско ниво со примена на математичка анализа која ќе даде резултати кои можат лесно компјутерски да се претстават и обработат, без посуштинска анализа на процесите кои се одвиваат на микроскопско ниво.

Увидувајќи дека протеините на микроскопско ниво се еден хаотичен систем од атоми со голем број на меѓуатомски дејствија и воден од идеата дека до одреден степен на тој хаотичен систем може да му се даде некоја просторно-логичка димензија, во овој труд ќе биде предложен нов пристап на моделирање на просторната структура на скелетот на протеините. Главната идеја која се проследува во рамки на моделот е да се најде формално-математички систем кој ќе ја опише просторната релација меѓу скелетните протеински атоми, при што моделот ќе биде изведен врз база на примена на концепти од векторска алгебра. По формирањето на теоретската концепција на моделот, истиот ќе биде применет на реални протеински структури а генералните резултати од таа примена на голем број на протеини од различни протеински фамилии ќе бидат дадени во делот за анализа на експериментално добиените податоци. Моделот кој го предложив го нареков скелетен модел, бидејќи истиот ја моделира преку математички формален систем просторната зависност меѓу скелетните атоми.

## 3.2 СКЕЛЕТЕН МОДЕЛ

Протеините се секвенци изградени од дваесет различни аминокиселини. Секоја аминокиселина е изградена од фиксен дел кој содржи централен јаглероден атом  $C_\alpha$ , амино-група  $NH_2$  и една една карбоксилна група  $COOH$ . Централниот јаглероден атом  $C_\alpha$  е поврзан со резидуум кој има различна структура за секоја од дваесетте аминокиселини. Кога повеќе аминокиселини меѓусебно ќе изреагираат со формирање на меѓусебни пептидни врски доаѓа до создавање на аминокиселински вериги кои можат да бидат пептиди или полипептиди кои понатаму градат конкретни протеини. На Слика 3.2.1 е прикажана општата структура на дел од протеин.



Слика 3.2.1 Општа структура на дел од протеин

Главната хоризонтална линија на Слика 3.2.1 го прикажува скелетот на протеинот кој е составен од низа на азотни атоми означени со  $N_i, i \in Z^+$ , централни јаглеродни атоми означени со  $A_i, i \in Z^+$  и низа на јаглеродни атоми означени со  $C_i, i \in Z^+$ . Секој јаглероден атом  $C_i, i \in Z^+$  во рамки на скелетот е поврзан со кислороден атом  $O_i, i \in Z^+$ , додека пак секој централен јаглероден атом  $A_i, i \in Z^+$  е поврзан со резидуум  $R_i, i \in Z^+$ .

Секој од атомите во рамки на протеинот има одредена просторна позиција која е одредена со координатната тројка  $(x, y, z)$  а координатите на сите атоми се дадени во координатната секција за секој протеин во рамки на Protein Data Bank базата на податоци.

Скелетот на протеините како што е прикажано на Слика 3.2.1 се состои од низа на натриумови, централни јаглеродни и јаглеродни атоми. Просторната распределба на тие атоми е определена со нивните просторни тродимензионални координати. Низата на  $n$  скелетни атоми ќе ја означиме со  $A_1, A_2, \dots, A_{n-1}, A_n$ . Ќе земеме дека елементите на позиции  $A_{3i-2}, i \in Z^+(A_1, A_4, A_7, \dots)$  се азотните атоми, елементите на позиции  $A_{3i-1}, i \in Z^+(A_2, A_5, A_8, \dots)$  се централните јаглеродни атоми и елементите на позиции  $A_{3i}, i \in Z^+(A_3, A_6, A_9, \dots)$  се јаглеродните атоми.

За низата на  $n$  скелетни атоми со стохастички распределени координатни положби ќе биде предложен математички модел кој ќе даде просторно-логички опис на скелетот на протеинот, односно низата на  $n$  скелетни атоми ќе добие математички дефинирана просторно-логичка димензија.

За секои два соседни атоми во рамки на протеинскиот скелет  $A_i$  и  $A_{i+1}$ ,  $i \in 1, 2, \dots, n-2, n-1$  може да

се определи позиционен вектор  $\overrightarrow{A_i A_{i+1}}$  кој дава информација за взаемната ориентација на двата соседни атоми и за растојанието меѓу тие два соседни атоми. Позициониот вектор меѓу два соседни

атоми со координати  $A_i(x_i, y_i, z_i)$  и  $A_{i+1}(x_{i+1}, y_{i+1}, z_{i+1})$  е векторот  $\overrightarrow{A_i A_{i+1}} = \begin{bmatrix} x_{i+1} - x_i \\ y_{i+1} - y_i \\ z_{i+1} - z_i \end{bmatrix}$ . Пред врзувањето

на двата соседни атоми во вектор, атомите  $A_i$  и  $A_{i+1}$  беа два атоми со стохастички распределени

координатни позиции. По определувањето на позиционен вектор  $\overrightarrow{A_i A_{i+1}}$  за истите, меѓу нив е воспоставена просторна релација за нивната релативна меѓусебна положба, нивната взаема просторна ориентација и за нивната одалеченост.

Секој позиционен вектор  $\overrightarrow{A_i A_{i+1}}$ ,  $i \in 1, 2, \dots, n-2, n-1$  дава информација за насоката на просторно локално скршнување на низата на скелетни атоми, индиректно го определува растојанието меѓу соседните атоми  $\left| \overrightarrow{A_i A_{i+1}} \right| = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}$ ,  $i \in 1, 2, \dots, n-2, n-1$  и дава информација за релативната позиција на атомот  $A_{i+1}$  во однос на атомот  $A_i$ .

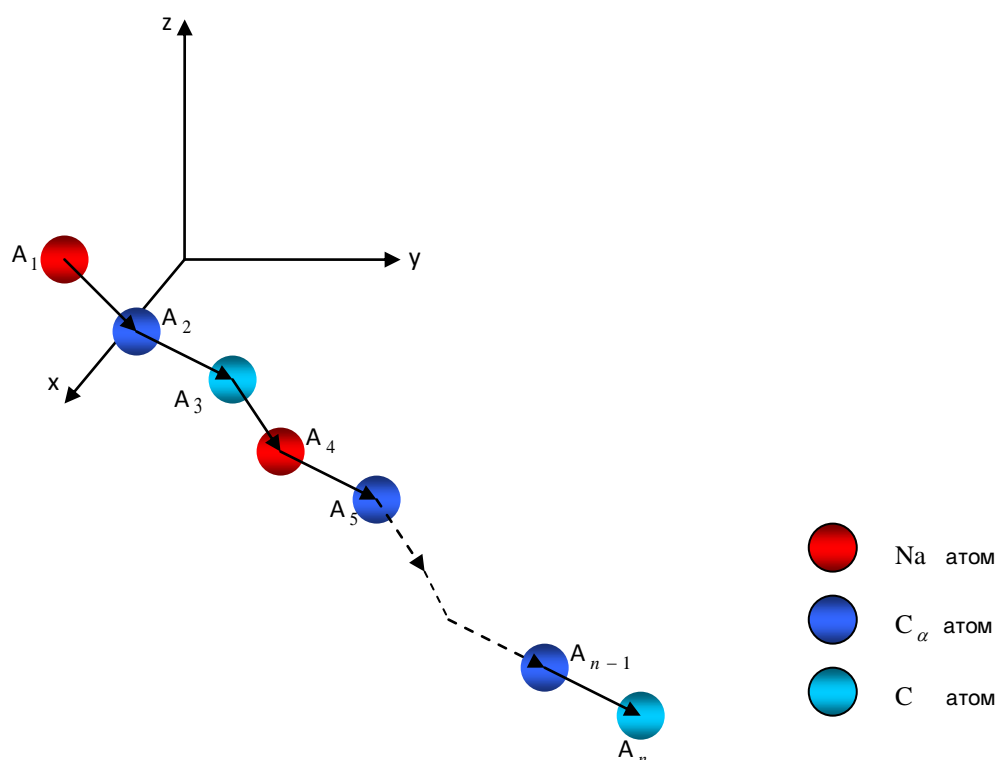
Ако  $x_{i+1} - x_i > 0$ , тогаш позицијата на атомот  $A_{i+1}$  во однос на  $x$ -оската е  $x_{i+1} - x_i$  единици во десно по  $x$ -оската во однос на позицијата на атомот  $A_i$  во однос на  $x$ -оската. Ако  $x_{i+1} - x_i < 0$ , тогаш позицијата на атомот  $A_{i+1}$  во однос на  $x$ -оската е  $x_{i+1} - x_i$  единици во лево по  $x$ -оската во однос на позицијата на атомот  $A_i$  во однос на  $x$ -оската.

Ако  $y_{i+1} - y_i > 0$ , тогаш позицијата на атомот  $A_{i+1}$  во однос на  $y$ -оската е  $y_{i+1} - y_i$  единици нагоре по  $y$  оската во однос на позицијата на атомот  $A_i$  во однос на  $y$ -оската. Ако  $y_{i+1} - y_i < 0$ , тогаш позицијата на атомот  $A_{i+1}$  во однос на  $y$ -оската е  $y_{i+1} - y_i$  единици надолу по  $y$ -оската во однос на позицијата на атомот  $A_i$  во однос на  $y$ -оската.

Ако  $z_{i+1} - z_i > 0$ , тогаш позицијата на атомот  $A_{i+1}$  во однос на  $z$ -оската е  $z_{i+1} - z_i$  единици нагоре по  $z$  оската во однос на позицијата на атомот  $A_i$  во однос на  $z$ -оската. Ако  $z_{i+1} - z_i < 0$ , тогаш позицијата на атомот  $A_{i+1}$  во однос на  $z$ -оската е  $z_{i+1} - z_i$  единици надолу по  $z$  оската во однос на позицијата на атомот  $A_i$  во однос на  $z$ -оската.



За низата на  $n$  скелетни атоми  $A_1, A_2, \dots, A_{n-1}, A_n$  можеме да ги определиме позиционите вектори за секои два соседни атоми. Такви позициони вектори за протеински скелет кој се состои од  $n$  атоми ќе има вкупно  $n - 1$  и истите се:  $\overrightarrow{A_1 A_2}, \overrightarrow{A_2 A_3}, \dots, \overrightarrow{A_{n-2} A_{n-1}}, \overrightarrow{A_{n-1} A_n}$ . Ланецот на соседни позициони вектори дава генерална слика за насоченоста на низата на скелетни атоми и секој од векторите во тој ланец може да се употреби за пресметување на растојанијата меѓу атомите кои тој вектор ги поврзува со формулата за модул на вектор  $\left| \overrightarrow{A_i A_{i+1}} \right| = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}, i \in 1, 2, \dots, n - 2, n - 1$ .

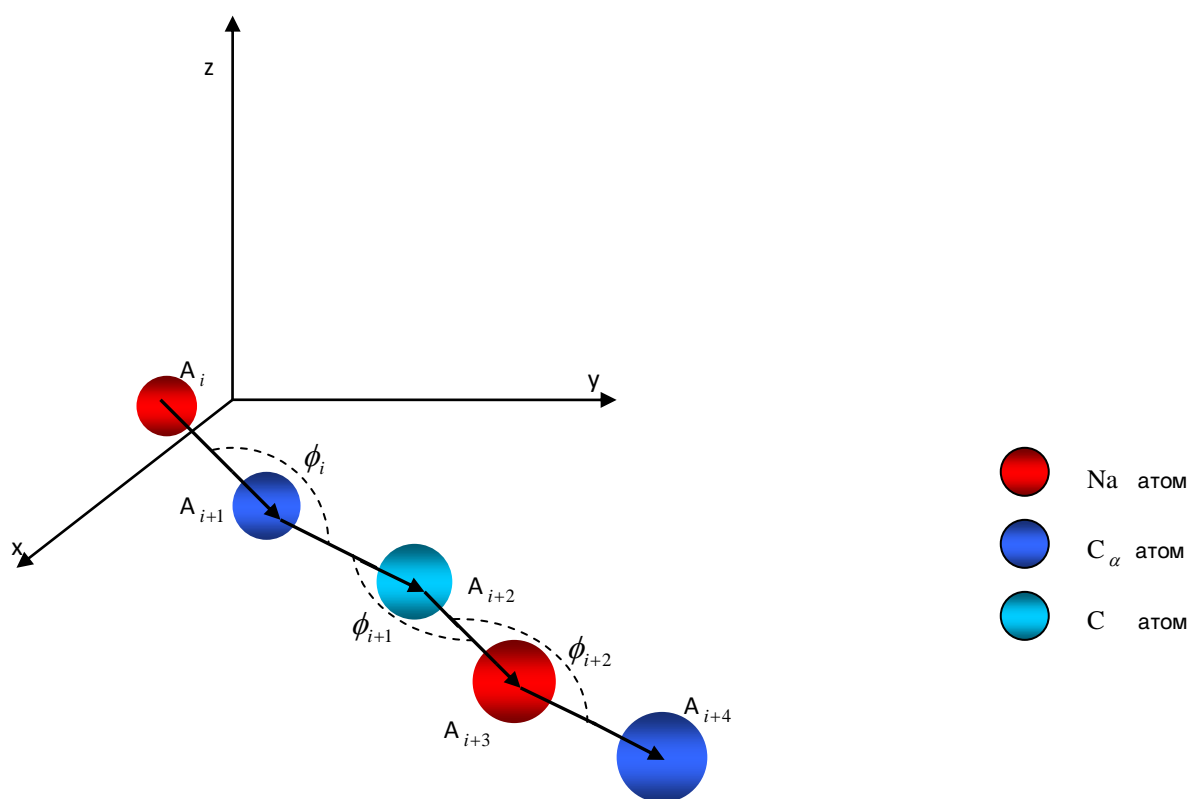


Слика 3.2.1 Низа на соседни скелетни позициони вектори

Низата на  $n - 1$  соседни позициони вектори  $\overrightarrow{A_1 A_2}, \overrightarrow{A_2 A_3}, \dots, \overrightarrow{A_{n-2} A_{n-1}}, \overrightarrow{A_{n-1} A_n}$  е првиот параметар во рамки на предложениот скелетен модел кој учествува во моделирањето на просторната структура на скелетот на протеинот.

Секои два соседни вектори од низата на  $n - 1$  позициони вектори меѓусебно формираат агол. Аглите меѓу секои два соседни позициони вектори во ланецот на позициони вектори ќе ги дефинираме како агли на свртување, бидејќи истите ја определуваат насоката на простирање на секој нареден позиционен вектор (почнувајќи од векторот  $\overrightarrow{A_2 A_3}$ ) изразена преку степени на свртување од тековната насока на претходниот позиционен вектор во рамки на ланецот на позициони вектори.

Кај ланецот на  $n - 1$  соседни позициони вектори такви агли ќе има вкупно  $n - 2$ . Со  $\phi_1$  ќе го означиме аголот на свртување меѓу првите два позициони вектори  $\overrightarrow{A_1 A_2}$  и  $\overrightarrow{A_2 A_3}$ , со  $\phi_2$  ќе го означиме аголот на свртување меѓу наредните два позициони вектори  $\overrightarrow{A_2 A_3}$  и  $\overrightarrow{A_3 A_4}$ , ..., со  $\phi_{n-2}$  ќе го означиме аголот на свртување меѓу предпоследниот и последниот позиционен вектор  $\overrightarrow{A_{n-2} A_{n-1}}$  и  $\overrightarrow{A_{n-1} A_n}$ .



Слика 3.2.2 Агли на свртување

Секои два соседни позициони вектори  $\overrightarrow{A_i A_{i+1}}$  и  $\overrightarrow{A_{i+1} A_{i+2}}$ ,  $i \in 1, 2, \dots, n-3, n-2$  формираат триаголник чии темиња се атомите  $A_i, A_{i+1}$  и  $A_{i+2}$ . Во рамки на низата на  $n-1$  соседни позициони вектори  $\overrightarrow{A_1 A_2}, \overrightarrow{A_2 A_3}, \dots, \overrightarrow{A_{n-2} A_{n-1}}, \overrightarrow{A_{n-1} A_n}$  можат да се формираат вкупно  $n-2$  соседни триаголници  $\Delta A_1 A_2 A_3, \Delta A_2 A_3 A_4, \dots, \Delta A_{n-3} A_{n-2} A_{n-1}, \Delta A_{n-2} A_{n-1} A_n$ . Аголот на свртување меѓу соседните позициони вектори  $\overrightarrow{A_i A_{i+1}}$  и  $\overrightarrow{A_{i+1} A_{i+2}}$  може да се определи со примена на косинусната теорема за триаголникот  $\Delta A_i A_{i+1} A_{i+2}$  на начинот даден во продолжение.

Векторот  $\overrightarrow{A_i A_{i+2}}$  е векторот спроти аголот на свртување  $\phi_i$  меѓу векторите  $\overrightarrow{A_i A_{i+1}}$  и  $\overrightarrow{A_{i+1} A_{i+2}}$  кај триаголникот  $\Delta A_i A_{i+1} A_{i+2}$ . Должините на страните  $A_i A_{i+1}, A_{i+1} A_{i+2}$  и  $A_i A_{i+2}$  во рамки на триаголникот  $\Delta A_i A_{i+1} A_{i+2}$  се модулите на векторите  $\left| \overrightarrow{A_i A_{i+1}} \right|, \left| \overrightarrow{A_{i+1} A_{i+2}} \right|$  и  $\left| \overrightarrow{A_i A_{i+2}} \right|$ . Векторот  $\overrightarrow{A_i A_{i+2}} = \overrightarrow{A_i A_{i+1}} + \overrightarrow{A_{i+1} A_{i+2}}$ .

Позиционите вектори  $\overrightarrow{A_i A_{i+1}} = \begin{bmatrix} x_{i+1} - x_i \\ y_{i+1} - y_i \\ z_{i+1} - z_i \end{bmatrix} = \begin{bmatrix} x_{i,i+1} \\ y_{i,i+1} \\ z_{i,i+1} \end{bmatrix}$  и  $\overrightarrow{A_{i+1} A_{i+2}} = \begin{bmatrix} x_{i+2} - x_{i+1} \\ y_{i+2} - y_{i+1} \\ z_{i+2} - z_{i+1} \end{bmatrix} = \begin{bmatrix} x_{i+1,i+2} \\ y_{i+1,i+2} \\ z_{i+1,i+2} \end{bmatrix}$  се веќе

познати, па векторот  $\overrightarrow{A_i A_{i+2}} = \overrightarrow{A_i A_{i+1}} + \overrightarrow{A_{i+1} A_{i+2}} = \begin{bmatrix} x_{i,i+1} + x_{i+1,i+2} \\ y_{i,i+1} + y_{i+1,i+2} \\ z_{i,i+1} + z_{i+1,i+2} \end{bmatrix}$ .

Квадратот од должината на страната спроти аголот на свртување  $\phi_i$  според косинусната теорема во рамки на триаголникот  $\Delta A_i A_{i+1} A_{i+2}$  е  $\left| \overrightarrow{A_i A_{i+2}} \right|^2 = \left| \overrightarrow{A_i A_{i+1}} \right|^2 + \left| \overrightarrow{A_{i+1} A_{i+2}} \right|^2 - 2 \left| \overrightarrow{A_i A_{i+1}} \right| \cdot \left| \overrightarrow{A_{i+1} A_{i+2}} \right| \cos \phi_i$ .

Бидејќи  $\left| \overrightarrow{A_i A_{i+2}} \right|^2 = (x_{i,i+1} + x_{i+1,i+2})^2 + (y_{i,i+1} + y_{i+1,i+2})^2 + (z_{i,i+1} + z_{i+1,i+2})^2$ ,  $\left| \overrightarrow{A_i A_{i+1}} \right|^2 = x_{i,i+1}^2 + y_{i,i+1}^2 + z_{i,i+1}^2$

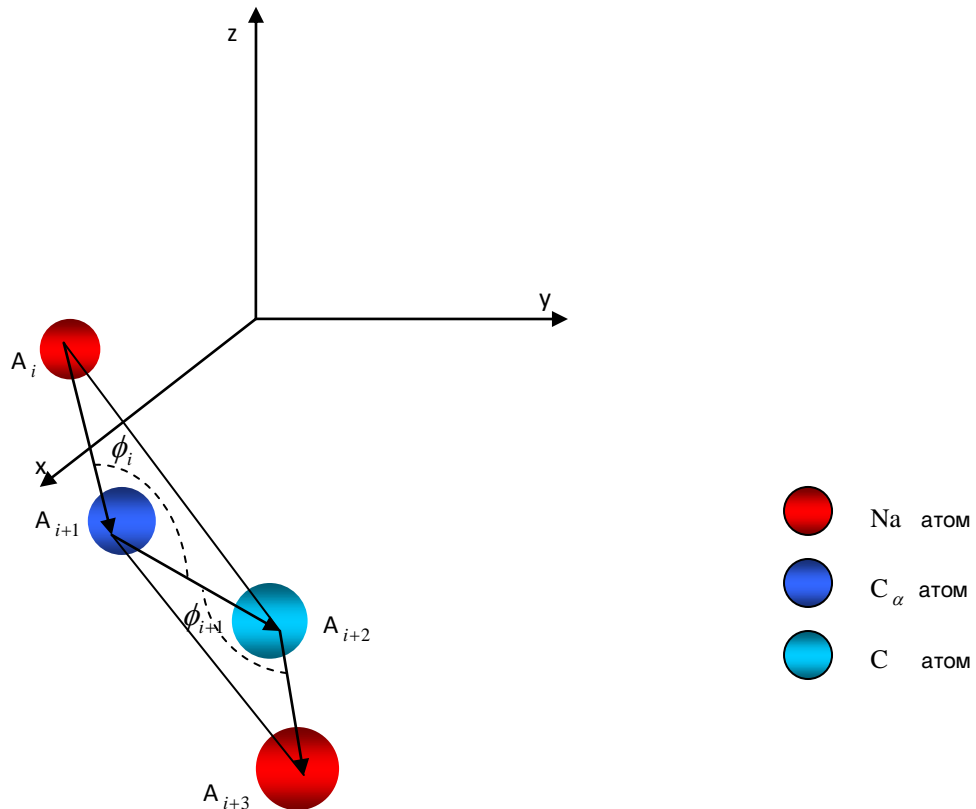
и  $\left| \overrightarrow{A_{i+1} A_{i+2}} \right|^2 = x_{i+1,i+2}^2 + y_{i+1,i+2}^2 + z_{i+1,i+2}^2$ , по замена во претходното равенство се добива дека

$$2x_{i,i+1} \cdot x_{i+1,i+2} + 2y_{i,i+1} \cdot y_{i+1,i+2} + 2z_{i,i+1} \cdot z_{i+1,i+2} = -2 \left| \overrightarrow{A_i A_{i+1}} \right| \cdot \left| \overrightarrow{A_{i+1} A_{i+2}} \right| \cos \phi_i.$$

Ако од претходното равенство се изрази аголот  $\phi_i$ , истиот е  $\phi_i = \cos^{-1} \left( - \frac{\vec{A_i A_{i+1}} \cdot \vec{A_{i+1} A_{i+2}}}{\left| \vec{A_i A_{i+1}} \right| \cdot \left| \vec{A_{i+1} A_{i+2}} \right|} \right)$ , каде

$\vec{A_i A_{i+1}} \cdot \vec{A_{i+1} A_{i+2}}$  е скаларниот производ меѓу соседните позициони вектори  $\vec{A_i A_{i+1}}$  и  $\vec{A_{i+1} A_{i+2}}$  и истиот

е  $\vec{A_i A_{i+1}} \cdot \vec{A_{i+1} A_{i+2}} = x_{i,i+1} \cdot x_{i+1,i+2} + y_{i,i+1} \cdot y_{i+1,i+2} + z_{i,i+1} \cdot z_{i+1,i+2}$ .



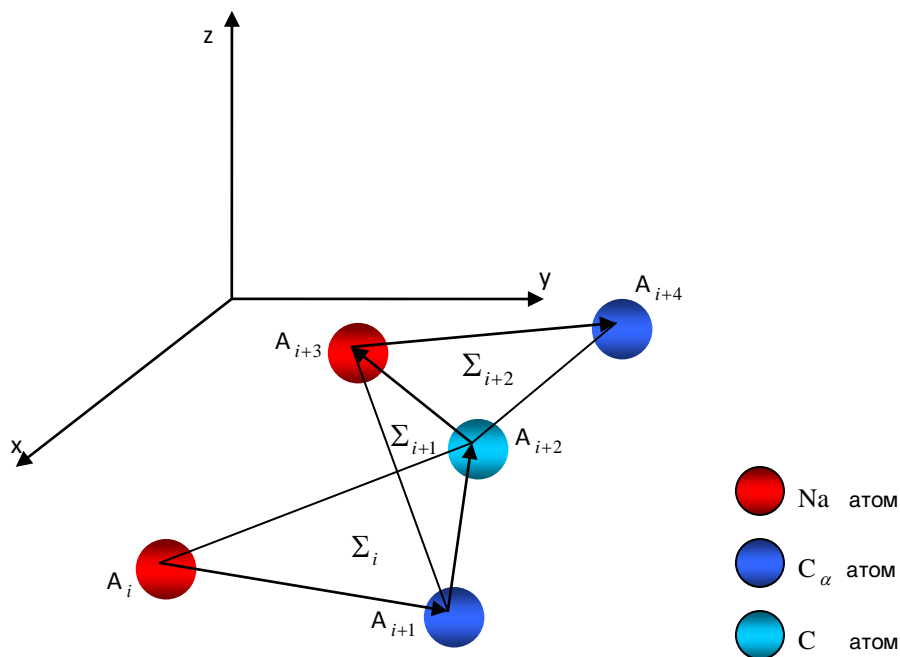
Слика 3.2.3 Пресметување на агли на свртување

По пресметувањата на скаларните производи меѓу секои два соседни позициони вектори и модулите на сите позициони вектори можеме да ја определиме низата на агли на свртување кои учествуваат во градбата на сликата за просторно-логичката структура на протеинскиот скелет преку равенството:

$\phi_i = \cos^{-1} \left( - \frac{\vec{A_i A_{i+1}} \cdot \vec{A_{i+1} A_{i+2}}}{\left| \vec{A_i A_{i+1}} \right| \cdot \left| \vec{A_{i+1} A_{i+2}} \right|} \right), i \in 1, 2, \dots, n-3, n-2$ . Низата на  $n-2$  агли на свртување

$\phi_1, \phi_2, \dots, \phi_{n-3}, \phi_{n-2}$  е вториот параметар во рамки на предложениот скелетен модел и истата ја дефинира искршеноста на протеинскиот скелет.

Низ центрите на секои три соседни скелетни атоми  $A_i, A_{i+1}, A_{i+2}, i \in 1, 2, \dots, n-3, n-2$  минува рамнина  $\Sigma_i, i \in 1, 2, \dots, n-3, n-2$  со определен наклон. За да ја најдеме равенката на скелетната рамнина  $\Sigma_i, i \in 1, 2, \dots, n-3, n-2$  мора да бидат познати координатите на еден атом од низата на три соседни атоми кои формираат рамнина и соседните позициони вектори  $\overrightarrow{A_i A_{i+1}}$  и  $\overrightarrow{A_{i+1} A_{i+2}}, i \in 1, 2, \dots, n-3, n-2$ .



Слика 3.2.4 Просторни скелетни рамнини  $\Sigma_i, \Sigma_{i+1}, \Sigma_{i+2}$

Векторскиот производ меѓу двата соседни позициони вектори  $\overrightarrow{A_i A_{i+1}}$  и  $\overrightarrow{A_{i+1} A_{i+2}}, i \in 1, 2, \dots, n-3, n-2$

ќе го означуваме со  $\overrightarrow{n_i}$  и истиот е вектор кој е нормален на рамнината која ја формираат трите неколинеарни атоми, односно двата соседни позициони вектори. Ако се познати двата соседни

позициони вектори во компонентна форма  $\overrightarrow{A_i A_{i+1}} = \begin{bmatrix} x_{i,i+1} \\ y_{i,i+1} \\ z_{i,i+1} \end{bmatrix}$  и  $\overrightarrow{A_{i+1} A_{i+2}} = \begin{bmatrix} x_{i+1,i+2} \\ y_{i+1,i+2} \\ z_{i+1,i+2} \end{bmatrix}, i \in 1, 2, \dots, n-3, n-2$

тогаш векторот кој е нормален на рамнината определена со положбите на трите соседни скелетни

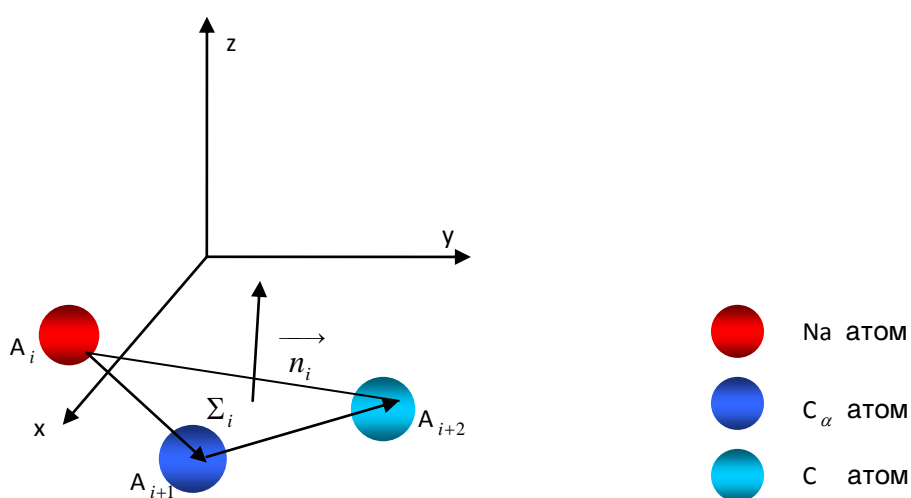
атоми и двата соседни позициони вектори  $\overrightarrow{A_i A_{i+1}}$  и  $\overrightarrow{A_{i+1} A_{i+2}}$  е нормалниот вектор  $\overrightarrow{n_i} = \begin{bmatrix} n_{x,i} \\ n_{y,i} \\ n_{z,i} \end{bmatrix}$ .

Првата компонента  $n_{x,i}$  во рамки на нормалниот вектор  $\vec{n_i}$  е

$$n_{x,i} = \begin{vmatrix} y_{i,i+1} & z_{i,i+1} \\ y_{i+1,i+2} & z_{i+1,i+2} \end{vmatrix} = y_{i,i+1} \cdot z_{i+1,i+2} - z_{i,i+1} \cdot y_{i+1,i+2}, \text{ втората компонента } n_{y,i} \text{ во рамки на нормалниот}$$

$$\text{вектор } \vec{n_i} \text{ е } n_{y,i} = \begin{vmatrix} z_{i,i+1} & x_{i,i+1} \\ z_{i+1,i+2} & x_{i+1,i+2} \end{vmatrix} = z_{i,i+1} \cdot x_{i+1,i+2} - x_{i,i+1} \cdot z_{i+1,i+2} \text{ и третата компонента } n_{z,i} \text{ во рамки на}$$

$$\text{нормалниот вектор } \vec{n_i} \text{ е } n_{z,i} = \begin{vmatrix} x_{i,i+1} & y_{i,i+1} \\ x_{i+1,i+2} & y_{i+1,i+2} \end{vmatrix} = x_{i,i+1} \cdot y_{i+1,i+2} - y_{i,i+1} \cdot x_{i+1,i+2}.$$



Слика 3.2.5 Приказ на рамнината  $\Sigma_i$  и нејзиниот нормален вектор  $\vec{n_i}$

Откако ќе биде определен нормалниот вектор на рамнината која ја формираат трите соседни атоми за да биде напишана равенката на рамнината потребно е да се знаат координатите на еден од трите соседни скелетни атоми. Нека координатите на скелетниот атом  $A_i$  во рамки на  $i$ -тата тројка на соседни скелетни атоми  $A_i, A_{i+1}, A_{i+2}$  се  $A_i(x_p, y_p, z_p)$ , тогаш равенката на рамнината која минува низ центрите на трите соседни скелетни атоми  $A_i, A_{i+1}$  и  $A_{i+2}$  е:

$$n_{x,i} \cdot x + n_{y,i} \cdot y + n_{z,i} \cdot z = n_{x,i} \cdot x_p + n_{y,i} \cdot y_p + n_{z,i} \cdot z_p.$$

Просторната ориентација на протеинскиот скелет е одредена со низата на скелетни рамнини

$\Sigma_1, \Sigma_2, \dots, \Sigma_{n-3}, \Sigma_{n-2}$  од облик  $n_{x,i} \cdot x + n_{y,i} \cdot y + n_{z,i} \cdot z = n_{x,i} \cdot x_p + n_{y,i} \cdot y_p + n_{z,i} \cdot z_p$ ,  $i \in 1, 2, \dots, n-3, n-2$ , кои се третиот параметар на предложениот скелетен модел.

Секои две соседни скелетни рамнини  $\Sigma_i$  и  $\Sigma_{i+1}$ ,  $i \in 1, 2, \dots, n-4, n-3$  меѓусебно формираат просторен агол кој ќе го нарекуваме торзионен агол. За низата на  $n-2$  соседни скелетни рамнини определени со положбите на секои три соседни скелетни атоми постојат вкупно  $n-3$  соседни торзиони агли  $\psi_1, \psi_2, \dots, \psi_{n-4}, \psi_{n-3}$ , каде  $\psi_i$  е торзиониот агол меѓу рамнините  $\Sigma_i$  и  $\Sigma_{i+1}$ ,  $i \in 1, 2, \dots, n-4, n-3$ .

Торзиониот аголот меѓу двете соседни рамнини  $\Sigma_i$  и  $\Sigma_{i+1}$ ,  $i \in 1, 2, \dots, n-4, n-3$  е аголот меѓу

нормалните вектори  $\vec{n}_i = \begin{bmatrix} n_{x,i} \\ n_{y,i} \\ n_{z,i} \end{bmatrix}$  и  $\vec{n}_{i+1} = \begin{bmatrix} n_{x,i+1} \\ n_{y,i+1} \\ n_{z,i+1} \end{bmatrix}$  на рамнините  $\Sigma_i$  и  $\Sigma_{i+1}$  соодветно. Двата нормални

вектори  $\vec{n}_i$  и  $\vec{n}_{i+1}$  можат да се доведат во состојба да имат заедничка почетна точка и аголот кои тие

ќе го образуваат е всушност аголот меѓу рамнините  $\Sigma_i$  и  $\Sigma_{i+1}$ . Бидејќи векторите  $\vec{n}_i$  и  $\vec{n}_{i+1}$  имаат

заедничка почетна точка во однос на нив може да се дефинира вектор  $\vec{\lambda}_i = \vec{n}_{i+1} - \vec{n}_i = \begin{bmatrix} n_{x,i+1} - n_{x,i} \\ n_{y,i+1} - n_{y,i} \\ n_{z,i+1} - n_{z,i} \end{bmatrix}$ .

Според косинусната теорема  $\left| \vec{\lambda}_i \right|^2 = \left| \vec{n}_{i+1} \right|^2 + \left| \vec{n}_i \right|^2 - 2 \left| \vec{n}_{i+1} \right| \cdot \left| \vec{n}_i \right| \cos \psi_i$ , каде  $\psi_i$  е аголот меѓу

нормалните вектори  $\vec{n}_i$  и  $\vec{n}_{i+1}$  и истиот е всушност аголот меѓу рамнините  $\Sigma_i$  и  $\Sigma_{i+1}$ .

Со замена за  $\left| \vec{\lambda}_i \right|^2 = (n_{x,i+1} - n_{x,i})^2 + (n_{y,i+1} - n_{y,i})^2 + (n_{z,i+1} - n_{z,i})^2 = n_{x,i+1}^2 - 2n_{x,i+1}n_{x,i} + n_{x,i}^2 + n_{y,i+1}^2 - 2n_{y,i+1}n_{y,i} + n_{y,i}^2 + n_{z,i+1}^2 - 2n_{z,i+1}n_{z,i} + n_{z,i}^2$ ,

$$\left| \vec{n}_i \right|^2 = n_{x,i}^2 + n_{y,i}^2 + n_{z,i}^2 \text{ и } \left| \vec{n}_{i+1} \right|^2 = n_{x,i+1}^2 + n_{y,i+1}^2 + n_{z,i+1}^2 \text{ во равенството } \left| \vec{\lambda}_i \right|^2 = \left| \vec{n}_{i+1} \right|^2 + \left| \vec{n}_i \right|^2 - 2 \left| \vec{n}_{i+1} \right| \left| \vec{n}_i \right| \cos \psi_i$$

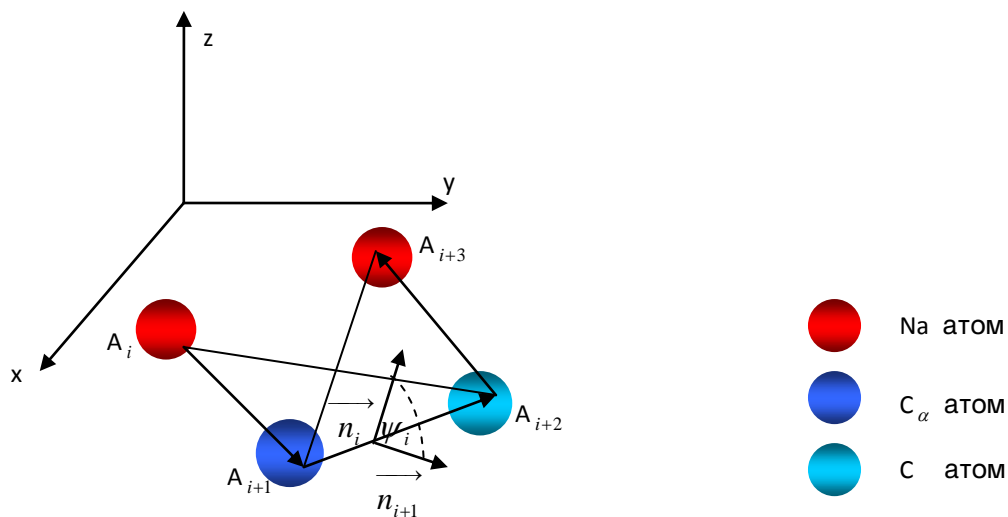
се добива дека  $-2n_{x,i+1}n_{x,i} - 2n_{y,i+1}n_{y,i} - 2n_{z,i+1}n_{z,i} = -2 \left| \vec{n}_{i+1} \right| \left| \vec{n}_i \right| \cos \psi_i$ , од каде ако се изрази аголот

$$\text{меѓу рамнините } \Sigma_i \text{ и } \Sigma_{i+1} \text{ за истиот се добива дека е } \psi_i = \cos^{-1} \left( \frac{\vec{n}_i \cdot \vec{n}_{i+1}}{\left| \vec{n}_i \right| \left| \vec{n}_{i+1} \right|} \right), i \in 1, 2, \dots, n-4, n-3.$$

Во претходното равенство  $\vec{n}_i \cdot \vec{n}_{i+1} = n_{x,i} \cdot n_{x,i+1} + n_{y,i} \cdot n_{y,i+1} + n_{z,i} \cdot n_{z,i+1}$  е скаларниот производ на

нормалните вектори  $\vec{n}_i$  и  $\vec{n}_{i+1}$  а  $\left| \vec{n}_i \right| = \sqrt{n_{x,i}^2 + n_{y,i}^2 + n_{z,i}^2}$  и  $\left| \vec{n}_{i+1} \right| = \sqrt{n_{x,i+1}^2 + n_{y,i+1}^2 + n_{z,i+1}^2}$  се модулите

на нормалните вектори  $\vec{n}_i$  и  $\vec{n}_{i+1}$ .



Слика 3.2.6 Приказ на торзионен агол  $\psi_i$

Доколку се примени претходниот концепт можат да се определат сите  $n-3$  торзиони агли  $\psi_1, \psi_2, \dots, \psi_{n-4}, \psi_{n-3}$  кои ја детализираат просторната структура на скелетот на протеинот. Низата на  $n-3$  торзиони агли  $\psi_1, \psi_2, \dots, \psi_{n-4}, \psi_{n-3}$  е четвртиот и последен параметар на предложениот скелетен модел.



### 3.3 Сублимат на карактеристиките на предложениот модел

Предложениот скелетен модел има за цел да даде математички дефинирана просторно-логичка димензија на протеинските скелетни атоми кон кои е пристапено како кон множество на честички со стохастички распределени координати во тродимензионалниот простор. Истиот го моделира скелетот на протеинот кој се состои од низа на азотни, централни јаглеродни и јаглеродни атоми и е опишан со математичкиот формален систем даден во продолжение.

$$\left\{ \begin{array}{l} \overrightarrow{A_1 A_2}, \overrightarrow{A_2 A_3}, \dots, \overrightarrow{A_{n-2} A_{n-1}}, \overrightarrow{A_{n-1} A_n} \\ \phi_1, \phi_2, \dots, \phi_{n-3}, \phi_{n-2} \\ \Sigma_1, \Sigma_2, \dots, \Sigma_{n-3}, \Sigma_{n-2} \\ \psi_1, \psi_2, \dots, \psi_{n-4}, \psi_{n-3} \end{array} \right.$$

- Низата на соседни позициони вектори  $\overrightarrow{A_1 A_2}, \overrightarrow{A_2 A_3}, \dots, \overrightarrow{A_{n-2} A_{n-1}}, \overrightarrow{A_{n-1} A_n}$  дава информација за релативната позиција на секој скелетен атом, меѓуатомските скелетни растојанија и генералната ориентираност и насоченост на протеинскиот скелет.
- Низата на агли на свртување  $\phi_1, \phi_2, \dots, \phi_{n-3}, \phi_{n-2}$  дава информација за дводимензионалната аглова ориентираност и насоченост на протеинскиот скелет.
- Низата на соседни скелетни рамнини  $\Sigma_1, \Sigma_2, \dots, \Sigma_{n-3}, \Sigma_{n-2}$  дава генерална претстава за просторната структура на протеинскиот скелет.
- Низата на торзиони агли:  $\psi_1, \psi_2, \dots, \psi_{n-4}, \psi_{n-3}$  ја детализира просторната структура, ориентација и насоченост на скелетот на протеинот.

# Просторно моделирање на конкретни протеини

Во овој дел на магистерскиот труд предложениот скелетен модел ќе се примени на конкретни протеини. За олеснување на математичките пресметки напишав C++ програма која како влез ги прима координатите на скелетните атоми во рамки на протеинот а на излез ги дава пресметаните параметри на скелетниот модел за конкретен протеин. Координатите на скелетните атоми се експериментално откриени и истите се исчитуваат од координатната секција на pdb датотеката за протеинот кој е предмет на моделирање.

Предложениот скелетен модел е применет на протеинот - Human Alpha-Defensin HNP-1. Истиот протеин има имуно-одбранбена функција и е откриен со NMR спектроскопија. HNP-1 протеинот е верига од 30 аминокиселини. Аминокиселините во рамки на HNP-1 протеинот подредени по редослед се: ALA, CYS, TYR, CYS, ARG, ILE, PRO, ALA, CYS, ILE, ALA, GLY, GLU, ARG, ARG, TYR, GLY, THR, CYS, THR, CYS, ILE, TYR, GLN, GLY, ARG, LEU, TRP, ALA, PHE, CYS, CYS. Секоја од овие 30 аминокиселини содржи 3 скелетни атоми (азотен, централен јаглероден и јаглероден атом) а низата од 30 аминокиселини формира скелет со  $3 \times 30 = 90$  атоми.

Според предложениот скелетен модел, скелетот на протеинот е опишан со математичко-формален систем составен од следниве параметри: позициони вектори за низата од скелетни атоми, низа на агли на свртување, низа на соседни скелетни рамнини и низа на торзиони агли кои ја детализираат сликата за просторната структура на скелетот на протеинот.

Кај протеинскиот скелет од деведесет атоми  $A_1, A_2, \dots, A_{89}, A_{90}$  елементите на позиции  $3i-2, i \in 1, 2, \dots, 29, 30$  се азотните атоми  $N_i$  од  $i$ -тата аминокиселина, елементите на позиции  $3i-1, i \in 1, 2, \dots, 29, 30$  се централните јаглеродни атоми  $A_i$  од  $i$ -тата аминокиселина и елементите на позиции  $3i, i \in 1, 2, \dots, 29, 30$  се јаглеродните атоми  $C_i$  од  $i$ -тата аминокиселина.

За скелетните атоми кои се деведесет на број можат да бидат определени осумдесет и девет позициони вектори  $\overrightarrow{A_1 A_2}, \overrightarrow{A_2 A_3}, \dots, \overrightarrow{A_{88} A_{89}}, \overrightarrow{A_{89} A_{90}}$  кои дават информација за просторната насоченост и ориентација на скелетот на протеинот, меѓуатомските растојанија и релативните координатни позиции на атомите  $A_{i+1}$  изразени во однос на координатните позиции на нивните претходници - атомите  $A_i$ .

Во рамки на низата од осумдесет и девет скелетни позициони вектори постојат поднизи на:

- $\overrightarrow{N_i A_i}$  - азотно-централнојаглеродни позициони вектори
- $\overrightarrow{A_i C_i}$  - централнојаглеродни-јаглеродни позициони вектори
- $\overrightarrow{C_i N_{i+1}}$  - јаглеродно-азотни позициони вектори

Приказот на позиционите вектори по групи според претходната класификација е даден во

продолжение, при што позиционите вектори  $\overrightarrow{A_i A_{i+1}} = \begin{bmatrix} x_{i,i+1} \\ y_{i,i+1} \\ z_{i,i+1} \end{bmatrix}$  се прикажени во облик

$$\overrightarrow{A_i A_{i+1}} = \begin{bmatrix} x_{i,i+1} & y_{i,i+1} & z_{i,i+1} \end{bmatrix}.$$

$\vec{N_i A_i}$ ПОЗИЦИОНИ ВЕКТОРИ
$\vec{N_1 A_1} = [-0.51 \ -1.356 \ -0.346]$
$\vec{N_2 A_2} = [0.667 \ -0.613 \ -1.156]$
$\vec{N_3 A_3} = [-0.872 \ -0.626 \ -0.973]$
$\vec{N_4 A_4} = [0.423 \ -1.315 \ -0.456]$
$\vec{N_5 A_5} = [1.166 \ 0.0249996 \ -0.894]$
$\vec{N_6 A_6} = [-0.349 \ 0.284 \ -1.367]$
$\vec{N_7 A_7} = [1.145 \ 0.587 \ -0.763]$
$\vec{N_8 A_8} = [0.102 \ 1.425 \ 0.282001]$
$\vec{N_9 A_9} = [1.099 \ 0.398 \ 0.873]$
$\vec{N_{10} A_{10}} = [0.194 \ -1.031 \ 1.01]$
$\vec{N_{11} A_{11}} = [1.269 \ 0.316 \ 0.643]$
$\vec{N_{12} A_{12}} = [-0.106 \ 0.599 \ 1.323]$
$\vec{N_{13} A_{13}} = [-0.120001 \ -1.081 \ 0.976]$
$\vec{N_{14} A_{14}} = [-0.395 \ -1.292 \ -0.55]$
$\vec{N_{15} A_{15}} = [-0.681 \ -0.835 \ 0.978]$
$\vec{N_{16} A_{16}} = [0.363 \ -1.415 \ -0.067]$
$\vec{N_{17} A_{17}} = [-1.245 \ -0.756 \ -0.000999998]$
$\vec{N_{18} A_{18}} = [-0.783 \ 0.0860004 \ -1.237]$
$\vec{N_{19} A_{19}} = [-1.452 \ -0.0970011 \ 0.19]$
$\vec{N_{20} A_{20}} = [-0.612 \ -0.942999 \ -1.015]$
$\vec{N_{21} A_{21}} = [-1.333 \ 0.485001 \ -0.342]$
$\vec{N_{22} A_{22}} = [-0.197 \ 0.0439987 \ -1.443]$
$\vec{N_{23} A_{23}} = [1.321 \ -0.209 \ -0.582]$
$\vec{N_{24} A_{24}} = [0.708 \ 1.28 \ 0.0949993]$
$\vec{N_{25} A_{25}} = [0.72 \ 0.191 \ 1.257]$
$\vec{N_{26} A_{26}} = [0.089 \ 1.394 \ 0.475]$
$\vec{N_{27} A_{27}} = [1.356 \ 0.477 \ 0.199]$
$\vec{N_{28} A_{28}} = [0.0090003 \ 1.233 \ 0.754]$
$\vec{N_{29} A_{29}} = [-0.105 \ 1.14 \ -0.904]$
$\vec{N_{30} A_{30}} = [-0.41 \ 1.229 \ 0.685]$

Табела 4.1 Листа на  $\vec{N_i A_i}$  позициони вектори

Секој од позиционите вектори дадени претходно ја определува взаемната насоченост на скелетните азотни атоми  $N_i$  со скелетните централни јаглеродни атоми  $A_i$  во рамки на аминокиселините  $i \in 1, 2, \dots, 29, 30$ .

Така на пример да го земеме позициониот вектор  $\overrightarrow{N_2 A_2} = [0.667 \ -0.613 \ -1.156]$ . Од истиот може да се утврди дека скелетниот централен јаглероден атом  $A_2$  во рамки на втората аминокиселина се наоѓа 0.667 ангстрема десно по  $x$ -оската, 0.613 ангстрема надолу по  $y$ -оската и 1.156 ангстрема надолу по  $z$ -оската во однос на  $x, y$  и  $z$  координатите на скелетниот азотен атом  $N_2$  во рамки на втората аминокиселина. Со тоа е определена взаемната насоченост на претходните атоми. Оваа анализа е аналогна за секој скелетен позиционен вектор  $\overrightarrow{N_i A_i}$  кај протеинот HNP-1.

Средната вредност на растојанијата меѓу скелетните азотни атоми  $N_i$  и скелетните централни

јаглеродни атоми  $A_i$  кај протеинот HNP-1 е  $\bar{x} = \frac{\sum_{i=1}^{30} \left| \overrightarrow{N_i A_i} \right|}{30} = 1.46305 \text{ \AA}$ , со стандардна девијација

$$\sigma = \sqrt{\frac{\sum_{i=1}^{30} \left( \left| \overrightarrow{N_i A_i} \right| - \bar{x} \right)^2}{30}} = 0.0146302 \text{ \AA}.$$

Стандардната девијација е мерка за дисперзираноста на податоците. Доколку истата за едно податочно множество е мал број тогаш податоците се распределени во тесен опсег околу средната вредност, во спротивност податоците битно се разликуваат од средната вредност на податочното множество. Стандардната девијација на растојанијата меѓу скелетните азотни атоми  $N_i$  и централни јаглеродни атоми  $A_i$  е  $\sigma = 0.0146302 \text{ \AA}$ , која е вредност блиска до нула, од каде следи заклучокот дека

сите скелетни растојанија  $\left| \overrightarrow{N_i A_i} \right|$  кај протеинот HNP-1 се распределени во тесен опсег околу средната вредност на податочното множество од 1.46 \AA.

$\vec{A_i C_i}$ ПОЗИЦИОНИ ВЕКТОРИ
$\vec{A_1 C_1} = [0.701 \ -0.504 \ -1.261]$
$\vec{A_2 C_2} = [-1.011 \ -0.7 \ -0.885]$
$\vec{A_3 C_3} = [0.551 \ -1.36 \ -0.363]$
$\vec{A_4 C_4} = [1.211 \ 0.129001 \ -0.935]$
$\vec{A_5 C_5} = [-0.452 \ 0.152 \ -1.421]$
$\vec{A_6 C_6} = [1.198 \ 0.654 \ -0.673]$
$\vec{A_7 C_7} = [0.066 \ 1.534 \ 0.115]$
$\vec{A_8 C_8} = [1.198 \ 0.28 \ 0.886999]$
$\vec{A_9 C_9} = [0.156 \ -0.988 \ 1.151]$
$\vec{A_{10} C_{10}} = [1.284 \ 0.258 \ 0.779]$
$\vec{A_{11} C_{11}} = [-0.233 \ 0.589 \ 1.386]$
$\vec{A_{12} C_{12}} = [-0.00399971 \ -1.076 \ 1.083]$
$\vec{A_{13} C_{13}} = [-0.497 \ -1.272 \ -0.67]$
$\vec{A_{14} C_{14}} = [-0.587 \ -0.875 \ 1.094]$
$\vec{A_{15} C_{15}} = [0.377 \ -1.465 \ -0.202]$
$\vec{A_{16} C_{16}} = [-1.248 \ -0.873 \ -0.082]$
$\vec{A_{17} C_{17}} = [-0.749 \ 0.21 \ -1.31]$
$\vec{A_{18} C_{18}} = [-1.486 \ -0.208 \ 0.295]$
$\vec{A_{19} C_{19}} = [-0.629 \ -0.810999 \ -1.124]$
$\vec{A_{20} C_{20}} = [-1.443 \ 0.532 \ -0.242]$
$\vec{A_{21} C_{21}} = [-0.124 \ 0.171 \ -1.509]$
$\vec{A_{22} C_{22}} = [1.337 \ -0.181 \ -0.717]$
$\vec{A_{23} C_{23}} = [0.799 \ 1.3 \ -0.0200005]$
$\vec{A_{24} C_{24}} = [0.592 \ 0.183001 \ 1.394]$
$\vec{A_{25} C_{25}} = [0.037 \ 1.477 \ 0.372]$
$\vec{A_{26} C_{26}} = [1.44 \ 0.393 \ 0.29]$
$\vec{A_{27} C_{27}} = [-0.0550001 \ 1.309 \ 0.749]$
$\vec{A_{28} C_{28}} = [0.0649996 \ 1.175 \ -0.967]$
$\vec{A_{29} C_{29}} = [-0.542 \ 1.218 \ 0.754]$
$\vec{A_{30} C_{30}} = [0.661 \ 1.215 \ -0.647]$

Табела 4.2 Листа на  $\vec{A_i C_i}$  позициони вектори

Аналогно на претходната дискусија секој од позиционите вектори дадени претходно ја определува взаемната насоченост на скелетните централни јаглородни атоми  $A_i$  со скелетните јаглородни атоми  $C_i$  во рамки на аминокиселините  $i \in 1, 2, \dots, 29, 30$ .

Така на пример да го земеме позициониот вектор  $\overrightarrow{A_8 C_8} = [1.198 \ 0.28 \ 0.886999]$ . Од истиот може да се утврди дека скелетниот јаглороден атом  $C_8$  во рамки на осмата аминокиселина се наоѓа 1.198 ангстреми десно по  $x$ -оската, 0.28 ангстреми нагоре по  $y$ -оската и 0.886999 ангстреми нагоре по  $z$ -оската во однос на  $x, y$  и  $z$  координатите на скелетниот централен јаглороден атом  $A_8$  во рамки на осмата аминокиселина. Со ова е определена взаемната насоченост на атомите  $A_8$  и  $C_8$ . Претходната анализа аналогно може да се спроведе за секој скелетен позиционен вектор  $\overrightarrow{A_i C_i}$  кај протеинот HNP-1.

Средната вредност на растојанијата меѓу скелетните централни јаглородни атоми  $A_i$  и јаглородни

атоми  $C_i$  кај протеинот HNP-1 е  $\bar{x} = \frac{\sum_{i=1}^{30} \left| \overrightarrow{A_i C_i} \right|}{30} = 1.52427$  Å. Стандардната девијација за истите е

$$\sigma = \sqrt{\frac{\sum_{i=1}^{30} \left( \left| \overrightarrow{A_i C_i} \right| - \bar{x} \right)^2}{30}} = 0.0096554 \text{ Å.}$$

Стандардната девијација и во овој случај има вредност близу нула,

од каде следи заклучокот дека растојанијата меѓу скелетните централни јаглородни и јаглородни атоми се распределени во тесен опсег околу средната вредност од 1.52 Å.

$\overrightarrow{C_i N_{i+1}}$ ПОЗИЦИОНИ ВЕКТОРИ
$\overrightarrow{C_1 N_2} = [0.274 \ -1.309 \ -0.043]$
$\overrightarrow{C_2 N_3} = [0.21 \ 0.123 \ -1.3]$
$\overrightarrow{C_3 N_4} = [-0.553 \ -0.577 \ -1.047]$
$\overrightarrow{C_4 N_5} = [0.44 \ -1.116 \ -0.591]$
$\overrightarrow{C_5 N_6} = [0.317 \ 1.145 \ -0.561]$
$\overrightarrow{C_6 N_7} = [0.289 \ -0.356 \ -1.23]$
$\overrightarrow{C_7 N_8} = [0.401 \ 0.465 \ 1.177]$
$\overrightarrow{C_8 N_9} = [0.013 \ 1.18 \ 0.605]$
$\overrightarrow{C_9 N_{10}} = [0.453 \ -1.203 \ -0.320001]$
$\overrightarrow{C_{10} N_{11}} = [0.900001 \ 0.773 \ -0.6]$
$\overrightarrow{C_{11} N_{12}} = [1.036 \ 0.609 \ 0.557]$
$\overrightarrow{C_{12} N_{13}} = [-0.981 \ -0.897 \ -0.069]$
$\overrightarrow{C_{13} N_{14}} = [0.416 \ -1.143 \ 0.537]$
$\overrightarrow{C_{14} N_{15}} = [-1.273 \ -0.334001 \ -0.154]$
$\overrightarrow{C_{15} N_{16}} = [0.239 \ -0.704 \ 1.102]$
$\overrightarrow{C_{16} N_{17}} = [-1.021 \ 0.361 \ 0.769]$
$\overrightarrow{C_{17} N_{18}} = [-0.844 \ -0.967 \ -0.363]$
$\overrightarrow{C_{18} N_{19}} = [-0.855 \ 0.749 \ -0.706]$
$\overrightarrow{C_{19} N_{20}} = [-0.639 \ -1.11 \ 0.35]$
$\overrightarrow{C_{20} N_{21}} = [-0.2 \ 0.813 \ -1.033]$
$\overrightarrow{C_{21} N_{22}} = [-0.729 \ -0.914 \ -0.625]$
$\overrightarrow{C_{22} N_{23}} = [1.011 \ -0.448999 \ 0.742]$
$\overrightarrow{C_{23} N_{24}} = [-0.212 \ 0.843 \ 1.014]$
$\overrightarrow{C_{24} N_{25}} = [1.318 \ 0.160999 \ 0.0440006]$
$\overrightarrow{C_{25} N_{26}} = [0.389 \ 0.288 \ 1.246]$
$\overrightarrow{C_{26} N_{27}} = [0.535 \ 0.869 \ -0.834]$
$\overrightarrow{C_{27} N_{28}} = [1.099 \ 0.375999 \ 0.626]$
$\overrightarrow{C_{28} N_{29}} = [-1.152 \ 0.634 \ -0.207]$
$\overrightarrow{C_{29} N_{30}} = [0.761 \ 1.092 \ -0.0169997]$

Табела 4.3 Листа на  $\overrightarrow{C_i N_{i+1}}$  позициони вектори



Листата на позициони вектори дадени претходно ја определува взаемната насоченост на скелетните јаглеродни атоми  $C_i$  со скелетните азотни атоми  $N_{i+1}$  во рамки на аминокиселините  $i \in 1, 2, \dots, 28, 29$ .

Така на пример да го земеме позициониот вектор  $\overrightarrow{C_9 N_{10}} = [0.453 \ -1.203 \ -0.320001]$ . Од истиот може да се утврди дека скелетниот азотен атом  $N_{10}$  во рамки на десеттата аминокиселина се наоѓа 0.453 ангстреми десно по  $x$ -оската, 1.203 ангстреми надолу по  $y$ -оската и 0.320001 ангстреми надолу по  $z$ -оската во однос на  $x, y$  и  $z$  координатите на скелетниот јаглероден атом  $C_9$  во рамки на деветтата аминокиселина. Со тоа е определена взаемната насоченост на претходните два атоми. Оваа анализа може да се спроведе за секој скелетен позиционен вектор  $\overrightarrow{C_i N_{i+1}}$  кај протеинот HNP-1.

Средната вредност на растојанијата меѓу скелетните јаглеродни атоми  $C_i$  и скелетните азотни

атоми  $N_{i+1}$  кај протеинот HNP-1 е  $\bar{x} = \frac{\sum_{i=1}^{29} \left| \overrightarrow{C_i N_{i+1}} \right|}{29} = 1.32776 \text{ \AA}$ . Стандардната девијација за истите е

$$\sigma = \sqrt{\frac{\sum_{i=1}^{29} \left( \left| \overrightarrow{C_i N_{i+1}} \right| - \bar{x} \right)^2}{29}} = 0.00670848 \text{ \AA}. \text{ И овде стандардната девијација има вредност близу нула која}$$

укажува на фактот дека растојанијата меѓу скелетните јаглеродни атоми  $C_i$  и скелетните азотни атоми

$N_{i+1}$  се распределени во тесен опсег околу средната вредност од 1.33 \AA.

Нареден параметар кај предложениот пристап за моделирање на скелетот на протеинот е низата на агли на свртување. Бидејќи овде скелетот е составен од деведесет атоми во однос на истиот можат да бидат определени вкупно осудесет и осум соседни агли на свртување  $\phi_1, \phi_2, \dots, \phi_{87}, \phi_{88}$ .

Секои три соседни скелетни атоми  $N_i, A_i, C_i$  (азотен, централен јаглороден и јаглороден атом) во рамки на  $i$ -та аминокиселина формираат агол на свртување  $\angle N_i A_i C_i$ , секои два соседни скелетни атоми  $A_i, C_i$  (централен јаглороден и јаглороден атом) од  $i$ -та аминокиселина со секој скелетен азотен атом  $N_{i+1}$  од  $i+1$ -та аминокиселина формираат агол на свртување  $\angle A_i C_i N_{i+1}$  и секој скелетен јаглороден атом  $C_i$  од  $i$ -та аминокиселина со соседниот скелетен азотен и скелетен централен јаглороден атом од  $i+1$ -та аминокиселина ( $N_{i+1}$  и  $A_{i+1}$ ) формира агол на свртување  $\angle C_i N_{i+1} A_{i+1}$ .

Листата на такви агли кај скелетот на протеинот HNP-1 е дадена во продолжение.

$\angle N_i A_i C_i$ АГЛИ НА СВРТУВАЊЕ	$\angle A_i C_i N_{i+1}$ АГЛИ НА СВРТУВАЊЕ	$\angle C_i N_{i+1} A_{i+1}$ АГЛИ НА СВРТУВАЊЕ
$\phi_1 = 109.563$	$\phi_7 = 116.3$	$\phi_3 = 121.777$
$\phi_4 = 110.461$	$\phi_5 = 115.166$	$\phi_6 = 121.626$
$\phi_7 = 109.307$	$\phi_8 = 115.593$	$\phi_9 = 121.54$
$\phi_{10} = 110.139$	$\phi_{11} = 117.292$	$\phi_{12} = 121.047$
$\phi_{13} = 109.829$	$\phi_{14} = 114.861$	$\phi_{15} = 121.268$
$\phi_{16} = 108.299$	$\phi_{17} = 118.109$	$\phi_{18} = 122.685$
$\phi_{19} = 112.685$	$\phi_{20} = 115.35$	$\phi_{21} = 122.386,$
$\phi_{22} = 110.44$	$\phi_{23} = 116.029$	$\phi_{24} = 121.543$
$\phi_{25} = 110.609$	$\phi_{26} = 116.17$	$\phi_{27} = 121.396$
$\phi_{28} = 110.3$	$\phi_{29} = 115.985$	$\phi_{30} = 121.094$
$\phi_{31} = 110.608$	$\phi_{32} = 116.142$	$\phi_{33} = 120.948$
$\phi_{34} = 110.78$	$\phi_{35} = 116.112$	$\phi_{36} = 121.628$
$\phi_{37} = 110.562$	$\phi_{38} = 116.023$	$\phi_{39} = 121.628$
$\phi_{40} = 110.079$	$\phi_{41} = 115.644$	$\phi_{42} = 121.073$
$\phi_{43} = 110.259$	$\phi_{44} = 116.299$	$\phi_{45} = 121.271$
$\phi_{46} = 110.682$	$\phi_{47} = 116.25$	$\phi_{48} = 121.037$
$\phi_{49} = 110.442$	$\phi_{50} = 116.432$	$\phi_{51} = 121.66$
$\phi_{52} = 110.373$	$\phi_{53} = 116.295$	$\phi_{54} = 121.795$
$\phi_{55} = 110.394$	$\phi_{56} = 116.722$	$\phi_{57} = 122.569$
$\phi_{58} = 105.422$	$\phi_{59} = 117.976$	$\phi_{60} = 121.515$
$\phi_{61} = 110.107$	$\phi_{62} = 115.739$	$\phi_{63} = 121.361,$
$\phi_{64} = 110.051$	$\phi_{65} = 116.276$	$\phi_{66} = 120.893$
$\phi_{67} = 110.938$	$\phi_{68} = 116.4$	$\phi_{69} = 121.581$
$\phi_{70} = 110.574$	$\phi_{71} = 115.454$	$\phi_{72} = 122.222$
$\phi_{73} = 110.41$	$\phi_{74} = 116.329$	$\phi_{75} = 121.415$
$\phi_{76} = 111.268$	$\phi_{77} = 115.732$	$\phi_{78} = 120.616$
$\phi_{79} = 108.609$	$\phi_{80} = 116.89$	$\phi_{81} = 119.722$
$\phi_{82} = 109.098$	$\phi_{83} = 115.418$	$\phi_{84} = 122.066$
$\phi_{85} = 109.991$	$\phi_{86} = 116.346$	$\phi_{87} = 121.47$
$\phi_{88} = 110.372$		

Табела 4.4 Листа на агли на свртување

Средната вредност на аглие на свртување  $\angle N_i A_i C_i$  кај протеинот HNP-1 е  $\bar{x} = \frac{\sum_{i=1}^{30} \phi_{3i-2}}{30} = 110.088^0$ .

Стандардната девијација за истите е  $\sigma = \sqrt{\frac{\sum_{i=1}^{30} (\phi_{3i-2} - \bar{x})^2}{30}} = 1.16359^0$ . Релативно малата вредност на стандардната девијација укажува на конвергентна распределба на протеинските скелетни агли на свртување  $\angle N_i A_i C_i$  околу средната вредност од 110 степени.

Средната вредност на аглие на свртување  $\angle A_i C_i N_{i+1}$  кај протеинот HNP-1 е

$$\bar{x} = \frac{\sum_{i=1}^{29} \phi_{3i-1}}{29} = 116.184^0. \text{ Стандардната девијација за истите е } \sigma = \sqrt{\frac{\sum_{i=1}^{29} (\phi_{3i-1} - \bar{x})^2}{29}} = 0.716328^0.$$

Стандардната девијација има вредност близу нула, која укажува на фактот дека сите агли на свртување  $\angle A_i C_i N_{i+1}$  кај протеинот HNP-1 се распределени во тесен опсег околу средната вредност од 116 степени.

Средната вредност на скелетните агли на свртување  $\angle C_i N_{i+1} A_{i+1}$  кај протеинот HNP-1 е

$$\bar{x} = \frac{\sum_{i=1}^{29} \phi_{3i}}{29} = 121.477^0. \text{ Стандардната девијација за истите е } \sigma = \sqrt{\frac{\sum_{i=1}^{29} (\phi_{3i} - \bar{x})^2}{29}} = 0.585216^0.$$

Стандардната девијација има мала вредност која укажува на конвергентна распределба на аглие на свртување  $\angle C_i N_{i+1} A_{i+1}$  околу средната вредност од 121 степени.

Нареден параметар кој ја определува просторната структура на скелетот на протеинот кај предложениот скелетен модел е низата на соседни скелетни рамнини. Секои три соседни неколинеарни скелетни атоми формираат рамнина со точно определен наклон. За скелетот на протеинот HNP-1 кој се состои од деведесет атоми можат да се определат осумдесет и осум соседни скелетни рамнини  $\Sigma_1, \Sigma_2, \dots, \Sigma_{87}, \Sigma_{88}$ .

Секои три соседни скелетни атоми  $N_i, A_i, C_i$  (азотен, централен јаглороден и јаглороден атом) во рамки на  $i$ -та аминокиселина формираат рамнина  $\Sigma_{3i-2}, i \in 1, 2, \dots, 29, 30$ , секои два соседни скелетни атоми  $A_i, C_i$  (централен јаглороден и јаглороден атом) од  $i$ -та аминокиселина со секој скелетен азотен атом  $N_{i+1}$  од  $i+1$ -та аминокиселина формираат рамнина  $\Sigma_{3i-1}, i \in 1, 2, \dots, 28, 29$  и секој скелетен јаглороден атом  $C_i$  од  $i$ -та аминокиселина со соседниот скелетен азотен и скелетен централен јаглороден атом од  $i+1$ -та аминокиселина ( $N_{i+1}$  и  $A_{i+1}$ ) формира рамнина  $\Sigma_{3i}, i \in 1, 2, \dots, 28, 29$ . Равенките на сите скелетни рамнини кај HNP-1 протеинот се дадени во продолжение.

$N_i A_i C_i$ СКЕЛЕТНИ РАМНИНИ
$\Sigma_1 : 1.53553x - 0.885656y + 1.2076z - 0.284512 = 0$
$\Sigma_4 : -0.266695x + 1.75901y - 1.08664z + 6.44495 = 0$
$\Sigma_7 : -1.09604x - 0.852659y + 1.53085z - 0.490043 = 0$
$\Sigma_{10} : 1.28835x - 0.156711y + 1.64703z + 10.8165 = 0$
$\Sigma_{13} : 0.100364x + 2.06097y + 0.188532z + 21.7452 = 0$
$\Sigma_{16} : 0.702885x - 1.87254y - 0.568478z - 22.7042 = 0$
$\Sigma_{19} : 1.23795x - 0.182033y + 1.71769z + 19.5598 = 0$
$\Sigma_{22} : 1.18501x + 0.247363y - 1.67859z - 25.4762 = 0$
$\Sigma_{25} : 1.32062x - 1.12876y - 1.1479z - 22.7626 = 0$
$\Sigma_{28} : 1.06373x + 1.14571y + 1.37386z + 25.62 = 0$
$\Sigma_{31} : 0.059249x - 1.90865y + 0.821068z - 1.62068 = 0$
$\Sigma_{34} : 2.07227x + 0.109506y + 0.116452z - 22.3632 = 0$
$\Sigma_{37} : 1.96574x - 0.565472y - 0.384616z - 23.7755 = 0$
$\Sigma_{40} : -1.8947x + 0.75498y - 0.412778z + 23.4978 = 0$
$\Sigma_{43} : 1.60144x + 0.231144y + 1.31246z - 6.81093 = 0$
$\Sigma_{46} : 0.0575389x + 0.113382y - 2.08282z - 0.160307 = 0$
$\Sigma_{49} : 0.990569x - 1.6302y - 0.827694z - 30.2165 = 0$
$\Sigma_{52} : 0.231926x + 2.06917y + 0.290661z + 35.4383 = 0$
$\Sigma_{55} : 0.263119x - 1.75156y + 1.11656z - 24.2069 = 0$
$\Sigma_{58} : 0.768185x + 1.31654y - 1.68633z + 19.3732 = 0$
$\Sigma_{61} : -0.673384x - 1.96909y - 0.167802z - 39.0957 = 0$
$\Sigma_{64} : -0.29273x - 2.07054y - 0.0231693z - 39.2468 = 0$
$\Sigma_{67} : 0.760779x - 0.438597y + 1.88429z + 14.8323 = 0$
$\Sigma_{70} : 1.76693x - 0.930712y - 0.628195z - 15.541 = 0$
$\Sigma_{73} : -1.78554x - 0.221331y + 1.05637z + 4.04779 = 0$
$\Sigma_{76} : 0.217585x + 0.65819y - 1.97238z - 1.93181 = 0$
$\Sigma_{79} : 0.0967824x - 1.02659y + 1.80124z - 1.20441 = 0$
$\Sigma_{82} : -2.07826x + 0.057713y - 0.0695691z + 10.4359 = 0$
$\Sigma_{85} : 1.96063x + 0.569138y + 0.48999z - 2.64346 = 0$
$\Sigma_{88} : -1.62744x + 0.187515y - 1.31052z + 1.08653 = 0$

Табела 4.5  $N_i A_i C_i$  скелетни рамнини

$A_i C_i N_{i+1}$ СКЕЛЕТНИ РАМНИНИ
$\Sigma_2 : -1.62898x - 0.315371y - 0.779513z - 2.8618 = 0$
$\Sigma_5 : 1.01886x - 1.50015y + 0.0226469z - 6.20686 = 0$
$\Sigma_8 : 1.21447x + 0.777636y - 1.07001z + 2.28026 = 0$
$\Sigma_{11} : -1.1197x + 0.304301y - 1.40824z - 7.76818 = 0$
$\Sigma_{14} : 1.54177x - 0.704029y - 0.565724z - 13.594 = 0$
$\Sigma_{17} : -1.04401x + 1.27904y - 0.615494z + 3.9541 = 0$
$\Sigma_{20} : 1.75204x - 0.0315671y - 0.584444z - 14.5978 = 0$
$\Sigma_{23} : -0.877259x - 0.71326y + 1.41z + 18.9486 = 0$
$\Sigma_{26} : 1.70081x + 0.571323y + 0.259896z - 6.62302 = 0$
$\Sigma_{29} : -0.756967x + 1.4715y + 0.760331z + 19.7985 = 0$
$\Sigma_{32} : -0.516001x + 1.56568y - 0.752101z + 5.5982 = 0$
$\Sigma_{35} : 1.0457x - 1.0627y - 1.05197z - 18.9211 = 0$
$\Sigma_{38} : -1.44887x - 0.0118314y + 1.09722z + 17.5002 = 0$
$\Sigma_{41} : 0.500147x - 1.48306y - 0.917816z - 21.1391 = 0$
$\Sigma_{44} : -1.75664x - 0.463732y + 0.0847262z + 7.4232 = 0$
$\Sigma_{47} : -0.641735x + 1.04343y - 1.34186z + 19.4598 = 0$
$\Sigma_{50} : -1.343x + 0.833753y + 0.901523z + 19.2099 = 0$
$\Sigma_{53} : -0.0741068x - 1.30134y - 1.29085z - 24.999 = 0$
$\Sigma_{56} : -1.53149x + 0.938386y + 0.179962z + 13.072 = 0$
$\Sigma_{59} : -0.35281x - 1.44222y - 1.06676z - 33.7181 = 0$
$\Sigma_{62} : -1.4861x + 1.02256y + 0.237995z + 9.3282 = 0$
$\Sigma_{65} : -0.456234x - 1.71694y - 0.417321z - 38.1697 = 0$
$\Sigma_{68} : 1.33506x - 0.805946y + 0.949157z + 0.288553 = 0$
$\Sigma_{71} : -0.216381x + 1.81124y - 0.145883z + 25.5567 = 0$
$\Sigma_{74} : 1.73321x + 0.0986061y - 0.563897z - 2.47206 = 0$
$\Sigma_{77} : -0.579772x + 1.35611y + 1.04111z + 20.8561 = 0$
$\Sigma_{80} : 0.53781x + 0.857581y - 1.45927z - 1.16135 = 0$
$\Sigma_{83} : 0.369853x + 1.12744y + 1.39481z + 10.0297 = 0$
$\Sigma_{86} : -0.844073x + 0.56458y - 1.51876z - 2.31659 = 0$

Табела 4.6  $A_i C_i N_{i+1}$  скелетни рамнини

$C_i N_{i+1} A_{i+1}$ СКЕЛЕТНИ РАМНИНИ
$\Sigma_3 : 1.48685x + 0.288063y + 0.705141z + 2.61316 = 0$
$\Sigma_6 : -0.933479x + 1.33793y - 0.0242039z + 5.4933 = 0$
$\Sigma_9 : -1.11369x - 0.695049y + 0.971266z - 2.01126 = 0$
$\Sigma_{12} : 1.01248x - 0.295746y + 1.31226z + 7.11943 = 0$
$\Sigma_{15} : -1.40589x + 0.629128y + 0.489633z + 12.0098 = 0$
$\Sigma_{18} : 0.993638x - 1.18784y + 0.577263z - 3.64684 = 0$
$\Sigma_{21} : -1.54609x + 0.00697178y + 0.523995z + 12.8794 = 0$
$\Sigma_{24} : 0.78935x + 0.653547y - 1.29165z - 17.2847 = 0$
$\Sigma_{27} : -1.54495x - 0.51961y - 0.233661z + 6.03704 = 0$
$\Sigma_{30} : 0.686639x - 1.3401y - 0.696536z - 18.0407 = 0$
$\Sigma_{33} : 0.472064x - 1.42967y + 0.685118z - 5.13184 = 0$
$\Sigma_{36} : -0.950061x + 0.965736y + 0.95282z + 17.1808 = 0$
$\Sigma_{39} : 1.32245x + 0.016685y - 0.988958z - 15.8927 = 0$
$\Sigma_{42} : -0.455243x + 1.34987y + 0.8355z + 19.241 = 0$
$\Sigma_{45} : 1.6065x + 0.416039y - 0.0826323z - 6.89753 = 0$
$\Sigma_{48} : 0.581003x - 0.958426y + 1.22132z - 17.8253 = 0$
$\Sigma_{51} : 1.2274x - 0.759799y - 0.829745z - 17.5297 = 0$
$\Sigma_{54} : 0.0738272x + 1.18756y + 1.17048z + 22.7899 = 0$
$\Sigma_{57} : 1.4567x - 0.862785y - 0.0767441z - 11.5338 = 0$
$\Sigma_{60} : 0.22296x + 1.30859y + 0.986729z + 30.2065 = 0$
$\Sigma_{63} : 1.3464x - 0.928821y - 0.212133z - 8.4637 = 0$
$\Sigma_{66} : 0.416395x + 1.56858y + 0.38183z + 34.8751 = 0$
$\Sigma_{69} : -1.21784x + 0.738052y - 0.868204z - 0.238932 = 0$
$\Sigma_{72} : 0.193972x - 1.62504y + 0.135818z - 22.8895 = 0$
$\Sigma_{75} : -1.60012x - 0.073881y + 0.516634z + 2.4826 = 0$
$\Sigma_{78} : 0.570749x - 1.23737y - 0.923169z - 18.9933 = 0$
$\Sigma_{81} : -0.488355x - 0.823012y + 1.35168z + 0.796172 = 0$
$\Sigma_{84} : -0.337156x - 1.01967y - 1.24671z - 8.99952 = 0$
$\Sigma_{87} : 0.768912x - 0.514315y + 1.38299z + 2.10799 = 0$

Табела 4.7  $C_i N_{i+1} A_{i+1}$  скелетни рамнини

Просторната слика за скелетот на протеинот формирана со множество на скелетни рамнини  $\Sigma_1, \Sigma_2, \dots, \Sigma_{87}, \Sigma_{88}$  се надополнува со последниот параметар во рамки на предложениот скелетен модел - низата на торзиони агли.

За скелетот на протеинот HNP-1 можат да се определат вкупно осумдесет и седум различни торзиони агли:  $\psi_1, \psi_2, \dots, \psi_{86}, \psi_{87}$ . Скелетните рамнини можат да бидат  $N_i A_i C_i, A_i C_i N_{i+1}$  или  $C_i N_{i+1} A_{i+1}$  рамнини, па следствено на тоа секој торзионен агол од низата на торзиони агли  $\psi_1, \psi_2, \dots, \psi_{86}, \psi_{87}$  припаѓа кон една од класите на торзиони агли дадени во продолжение.

Класа на торзиони агли меѓу рамнините  $N_i A_i C_i$  и  $A_i C_i N_{i+1}$ :  $\psi_1, \psi_4, \dots, \psi_{82}, \psi_{85}$ .

Класа на торзиони агли меѓу рамнините  $A_i C_i N_{i+1}$  и  $C_i N_{i+1} A_{i+1}$ :  $\psi_2, \psi_5, \dots, \psi_{83}, \psi_{86}$ .

Класа на торзиони агли меѓу рамнините  $C_i N_{i+1} A_{i+1}$  и  $N_{i+1} A_{i+1} C_{i+1}$ :  $\psi_3, \psi_6, \dots, \psi_{84}, \psi_{87}$ .



ТОРЗИОНИ АГЛИ МЕЃУ РАМНИНТЕ $N_i A_i C_i$ и $A_i C_i N_{i+1}$	ТОРЗИОНИ АГЛИ МЕЃУ РАМНИНТЕ $A_i C_i N_{i+1}$ и $C_i N_{i+1} A_{i+1}$	ТОРЗИОНИ АГЛИ МЕЃУ РАМНИНТЕ $C_i N_{i+1} A_{i+1}$ и $N_{i+1} A_{i+1} C_{i+1}$
$\psi_1 = 143.563$	$\psi_2 = 179.802$	$\psi_3 = 100.858$
$\psi_4 = 140.925$	$\psi_5 = 179.267$	$\psi_6 = 92.6297$
$\psi_7 = 168.138$	$\psi_8 = 179.461$	$\psi_9 = 85.4139$
$\psi_{10} = 174.667$	$\psi_{11} = 179.027$	$\psi_{12} = 94.2826$
$\psi_{13} = 112.267$	$\psi_{14} = 179.079$	$\psi_{15} = 136.671$
$\psi_{16} = 139.331$	$\psi_{17} = 179.354$	$\psi_{18} = 46.0442$
$\psi_{19} = 72.6476$	$\psi_{20} = 179.216$	$\psi_{21} = 143.333$
$\psi_{22} = 163.31$	$\psi_{23} = 179.566$	$\psi_{24} = 58.6261$
$\psi_{25} = 69.8095$	$\psi_{26} = 179.914$	$\psi_{27} = 77.7515$
$\psi_{28} = 21.1153$	$\psi_{29} = 179.831$	$\psi_{30} = 54.0091$
$\psi_{31} = 164.905$	$\psi_{32} = 179.934$	$\psi_{33} = 74.799$
$\psi_{34} = 59.4427$	$\psi_{35} = 179.914$	$\psi_{36} = 143.756$
$\psi_{37} = 149.623$	$\psi_{38} = 179.597$	$\psi_{39} = 127.349$
$\psi_{40} = 116.564$	$\psi_{41} = 180$	$\psi_{42} = 78.6095$
$\psi_{43} = 137.845$	$\psi_{44} = 179.676$	$\psi_{45} = 84.8418$
$\psi_{46} = 40.6562$	$\psi_{47} = 179.71$	$\psi_{48} = 70.9142$
$\psi_{49} = 155.225$	$\psi_{50} = 179.796$	$\psi_{51} = 126.824$
$\psi_{52} = 142.282$	$\psi_{53} = 179.715$	$\psi_{54} = 102.456$
$\psi_{55} = 119.231$	$\psi_{56} = 176.758$	$\psi_{57} = 88.326$
$\psi_{58} = 95.1198$	$\psi_{59} = 176.58$	$\psi_{60} = 146.888$
$\psi_{61} = 106.088$	$\psi_{62} = 179.857$	$\psi_{63} = 63.5952$
$\psi_{64} = 14.2901$	$\psi_{65} = 179.972$	$\psi_{66} = 84.232$
$\psi_{67} = 33.6938$	$\psi_{68} = 179.905$	$\psi_{69} = 131.057$
$\psi_{70} = 121.057$	$\psi_{71} = 179.828$	$\psi_{72} = 87.3769$
$\psi_{73} = 167.103$	$\psi_{74} = 179.405$	$\psi_{75} = 113.726$
$\psi_{76} = 109.938$	$\psi_{77} = 178.275$	$\psi_{78} = 95.6672$
$\psi_{79} = 159.686$	$\psi_{80} = 179.02$	$\psi_{81} = 75.3147$
$\psi_{82} = 102.132$	$\psi_{83} = 179.634$	$\psi_{84} = 122.421$
$\psi_{85} = 122.798$	$\psi_{86} = 179.98$	$\psi_{87} = 154.871$

Табела 4.8 Листа на торзиони агли

Средната вредност на торзионите агли меѓу рамнините  $N_iA_iC_i$  и  $A_iC_iN_{i+1}$  е

$$\bar{x} = \frac{\sum_{i=1}^{29} \psi_{3i-2}}{29} = 114.602^{\circ}. \text{ Стандардната девијација за истите е } \sigma = \sqrt{\frac{\sum_{i=1}^{29} (\psi_{3i-2} - \bar{x})^2}{29}} = 45.8206^{\circ}. \text{ Во овој}$$

случај стандардната девијација има голема вредност која укажува на постоење на широк опсег на распределба на торзионите агли меѓу рамнините  $N_iA_iC_i$  и  $A_iC_iN_{i+1}$ .

$$\text{Средната вредност на торзионите агли меѓу рамнините } A_iC_iN_{i+1} \text{ и } C_iN_{i+1}A_{i+1} \text{ е } \bar{x} = \frac{\sum_{i=1}^{29} \psi_{3i-1}}{29} = 179.382^{\circ}.$$

$$\text{Стандардната девијација за истите е } \sigma = \sqrt{\frac{\sum_{i=1}^{29} (\psi_{3i-1} - \bar{x})^2}{29}} = 0.830625^{\circ}. \text{ Малата вредност на}$$

стандардната девијација укажува на конвергентна распределба на торзионите агли меѓу рамнините  $A_iC_iN_{i+1}$  и  $C_iN_{i+1}A_{i+1}$  околу  $180$  степени. Поради тоа што ковалентните пептидини врски меѓу аминокиселините се воспоставени меѓу скелетните јаглеродни атоми  $C_i$  и скелетните азотни атоми  $N_{i+1}$ , се доаѓа до заклучокот дека скелетните централни јаглеродни атоми и јаглеродни атоми од  $i$ -те аминокиселини се речиси компланарни со скелетните азотни и централни јаглеродни атоми од  $i+1$ -те аминокиселини (лежат речиси во една иста рамнина).

$$\text{Средната вредност на торзионите агли меѓу рамнините } C_iN_{i+1}A_{i+1} \text{ и } N_{i+1}A_{i+1}C_{i+1} \text{ е } \bar{x} = \frac{\sum_{i=1}^{29} \psi_{3i}}{29} = 98.7119^{\circ}.$$

$$\text{Стандардната девијација за истите е } \sigma = \sqrt{\frac{\sum_{i=1}^{29} (\psi_{3i} - \bar{x})^2}{29}} = 29.7271^{\circ}. \text{ Стандардната девијација има}$$

голема вредност која укажува на релативно широка дисперзија на торзионите агли меѓу рамнините  $C_iN_{i+1}A_{i+1}$  и  $N_{i+1}A_{i+1}C_{i+1}$ .

Со моделирање на скелетот на протеинот HNP-1 утврдивме значајни реални физички особености на истиот дадени во продолжение.

1. Растојанијата меѓу скелетните азотни и централни јаглеродни атоми  $N_i - A_i$  се релативно константни со мал опсег на варијација и средна вредност од 1.46305 Å.
2. Растојанијата меѓу скелетните централни јаглеродни и јаглеродни атоми  $A_i - C_i$  се релативно константни со мал опсег на варијација и средна вредност од 1.52427 Å.
3. Растојанијата меѓу скелетните јаглеродни и азотни атоми  $C_i - N_{i+1}$  се релативно константни со мал опсег на варијација и средна вредност од 1.32776 Å.
4. Аглите на свртување  $\angle N_i A_i C_i$  се релативно константни со мал опсег на варијација и средна вредност од  $110.088^\circ$ .
5. Аглите на свртување  $\angle A_i C_i N_{i+1}$  се релативно константни со мал опсег на варијација и средна вредност од  $116.184^\circ$ .
6. Аглите на свртување  $\angle C_i N_{i+1} A_{i+1}$  се релативно константни со мал опсег на варијација и средна вредност од  $121.477^\circ$ .
7. Скелетните централни јаглеродни атоми  $A_i$  и јаглеродни атоми  $C_i$  од  $i$ -те аминокиселини се речиси компланарни со скелетните азотни атоми  $N_{i+1}$  и централни јаглеродни атом  $A_{i+1}$  од  $i+1$ -те аминокиселини.

# Анализа на експериментално добиени податоци и примена на моделот

## 5.1 Анализа на експериментално добиени податоци

Примената на предложениот скелетен модел на протеини кои припаѓаат на различни протеински фамилии даде резултати интересни за анализа и дискусија. Резултатите покажаа константност на дел од параметрите на предложениот модел и менливост на останатите. Се утврди дека скелетните меѓуатомски растојанија и агли на свртување се релативно константни и дека менливост пројавуваат две од трите класи на торзиони агли. Класата на константни торзиони агли укажува на генерална скоро перфектна компланарност на скелетните централни јаглеродни атоми и јаглеродни атоми со скелетните азотни и централни јаглеродни атоми кај секои две соседни аминокиселини кај протеините.

Генералните резултати од примената на моделот на голем број на протеини од различни протеински фамилии ги даде резултатите дадени во продолжение, при што треба да се напомене дека векторските, аглебарските и статистичките пресметки беа направени со помош на самостојно напишан C++ програмски код.

- Растојанијата меѓу скелетните азотни и централни јаглеродни атоми  $N_i - A_i$  пројавуваат константност со средна вредност од 1.46215 Å и стандардна девијација од 0.01325 Å.
- Растојанијата меѓу скелетните централни јаглеродни и јаглеродни атоми  $A_i - C_i$  пројавуваат константност со средна вредност од 1.52468 Å и стандардна девијација од 0.01198 Å.
- Растојанијата меѓу скелетните јаглеродни и азотни атоми  $C_i - N_{i+1}$  пројавуваат константност со средна вредност од 1.33509 Å и стандардна девијација од 0.01455 Å.

- Аглите на свртување  $\angle N_i A_i C_i$  пројавуваат константност со средна вредност од  $110.62^\circ$  и стандардна девијација од  $1.169^\circ$ .
- Аглите на свртување  $\angle A_i C_i N_{i+1}$  пројавуваат константност со средна вредност од  $116.62^\circ$  и стандардна девијација од  $0.719^\circ$ .
- Аглите на свртување  $\angle C_i N_{i+1} A_{i+1}$  пројавуваат константност со средна вредност од  $121.08^\circ$  и стандардна девијација од  $0.5833^\circ$ .
- Средната вредност на аглите меѓу рамнините  $\Sigma A_i C_i N_{i+1}$  и  $\Sigma C_i N_{i+1} A_{i+1}$  е  $180.3^\circ$  со стандардна девијација од  $0.835^\circ$  што укажува на постоење на компланарност меѓу атомите  $A_i, C_i, N_{i+1}$  и  $A_{i+1}$ . Значи кај секој протеински скелет централните јаглеродни атоми од две соседни аминокиселини  $A_i, A_{i+1}$  со скелетниот јаглероден атом од првата аминокиселина  $C_i$  и скелетниот азотен атом  $N_{i+1}$  од наредната аминокиселина приближно лежат во иста рамнина.

## 5.2 Примена на предложениот скелетен модел

Предложениот скелетен модел може да се примени за компјутерска визуелизација и структурно просторно споредување на протеините. Секоја софтверска апликација за компјутерска визуелизација на протеинските молекули се базира на воспоставување на просторно-логичка релација помеѓу протеинските атоми. Процесот на воспоставување на просторно-логичка зависност помеѓу протеинските атоми се одвива во повеќе чекори.

Прво се одредува доменот на пресликувањето, од реална физичка структура во математичко-софтверска структура која во целост комплетно и точно ќе ја претставува реалната физичка структура на бараното ниво на разгледување, водејќи сметка таквата структура да биде погодна за софтверска обработка која на крајот ќе треба да генерира резултати кои ќе укажуваат на реални физички особености и релации кај протеините.

Несомнено дека предложениот скелетен модел воспоставува таква просторно-логичка релација помеѓу протеинските атоми, и истиот претставува своевидна база за компјутерска визуелизација на протеинската молекула од повеќе аспекти. Низата на соседни позициони вектори и низата на агли на свртување ја одредуваат дводимензионалната просторна структурираност на протеинската молекула, додека пак низата на соседни скелетни рамнини и низата на просторни торзиони агли ја одредуваат тродимензионалната просторна структурираност на протеинската молекула и истите можат да бидат софтверски имплементирани со примена на OpenGL или Direct3D развојни програмски визуелизациски околии.

Предложениот скелетен модел е посебно погоден за споредување на просторните структури на различни протеини. Целта на секое просторно споредување на протеини е да се идентификува максималното просторно совпаѓање помеѓу два протеини. Просторната респределба на секоја протеинска молекула во најголем дел е одредена со просторната распределба на скелетот на истата. Претходно утврдивме дека различната просторна структура на протеинските молекули а со тоа и на протеинските скелети се должи на широката дисперзија на аглите помеѓу рамнините  $\psi'$ :  $N_i A_i C_i$  и  $A_i C_i N_{i+1}$  и аглите помеѓу рамнините  $\psi''$ :  $C_i N_{i+1} A_{i+1}$  и  $N_{i+1} A_{i+1} C_{i+1}$ . Нека споредуваните протеини се протеините X и Y, со должини(број на аминокиселини)  $n$  и  $m$  соодветно. Примарната структура на протеинот X ќе ја означиме со  $X_1 X_2 \dots X_{n-1} X_n$ , а на протеинот Y со  $Y_1 Y_2 \dots Y_{m-1} Y_m$ , каде  $X_i, i \in 1, 2, \dots, n-1, n$  и  $Y_j, j \in 1, 2, \dots, m-1, m$  се аминокиселини од множеството на дваесет градебени протеински аминокиселини. Разгледуваме случај кога должината на протеинот X е помала од должината на протеинот Y,  $n < m$ . Во конкретниов случај сакаме да ја определиме подсеквенцата од протеинот Y која е со најголем процент на просторна сличност со протеинот X. Во рамки на протеинот Y такви подсеквенци има вкупно  $m - n + 1$ , и истите се:  $Y_1 Y_2 \dots Y_{n-1} Y_n, Y_2 Y_3 \dots Y_n Y_{n+1}, \dots, Y_{m-n+1} Y_{m-n+2} \dots Y_{m-1} Y_m$ . Секоја од претходните подсеквенци просторно може да се спореди со протеинот X, и за истите може да се определи степенот на просторно совпаѓање.

Како метрика за степен на просторно совпаѓање овде ќе ја користиме средната квадратна девијација, која може да се пресмета за секоја од претходните подсеквенци кај протеинот Y и протеинот

$$X \text{ по равенството: } d_i = \sqrt{\frac{\sum_{i=1}^n (\psi'_{X,i} - \psi'_{Y,i})^2 + \sum_{i=1}^n (\psi''_{X,i} - \psi''_{Y,i})^2}{n}}.$$

Величините во претходното равенство се следниве:

- $\psi'_{X,i}$  - торзиони агли помеѓу рамнините  $N_i A_i C_i$  и  $A_i C_i N_{i+1}$  во рамки на протеинот X.
- $\psi'_{Y,i}$  - торзиони агли помеѓу рамнините  $N_i A_i C_i$  и  $A_i C_i N_{i+1}$  во рамки на споредуваната подсеквенца од протеинот Y.
- $\psi''_{X,i}$  - торзиони агли помеѓу рамнините  $C_i N_{i+1} A_{i+1}$  и  $N_{i+1} A_{i+1} C_{i+1}$  во рамки на протеинот X.
- $\psi''_{Y,i}$  - торзиони агли помеѓу рамнините  $C_i N_{i+1} A_{i+1}$  и  $N_{i+1} A_{i+1} C_{i+1}$  во рамки на споредуваната подсеквенца од протеинот Y.

Така  $d_1$  го претставува степенот на просторно совпаѓање помеѓу протеинските секвенци  $Y_1 Y_2 \dots Y_{n-1} Y_n$  и  $X_1 X_2 \dots X_{n-1} X_n$ ,  $d_2$  го претставува степенот на просторно совпаѓање помеѓу протеинските секвенци  $Y_2 Y_3 \dots Y_n Y_{n+1}$  и  $X_1 X_2 \dots X_{n-1} X_n$ , ...,  $d_{m-n+1}$  го претставува степенот на просторно совпаѓање помеѓу протеинските секвенци  $Y_{m-n+1} Y_{m-n+2} \dots Y_{m-1} Y_m$  и  $X_1 X_2 \dots X_{n-1} X_n$ , при што сите вредности  $d_1, d_2, \dots, d_{m-n}, d_{m-n+1}$  се пресметуваат по претходно утврдената метрика. Најмалата вредност  $d_i$  од елементите  $d_1, d_2, \dots, d_{m-n}, d_{m-n+1}$ , ја одредува поднизата од протеинот Y со најголем процент на просторна структурна сличност со протеинот X. Нека таа вредност е елементот на позиција  $k$  во рамки на низата  $d_1, d_2, \dots, d_{m-n}, d_{m-n+1}$ , односно  $d_k = \min\{d_1, d_2, \dots, d_{m-n}, d_{m-n+1}\}$ . Тогаш со директно користење на еден од параметрите на предложениот модел (низата на просторни агли) и дефинирање на соодветна метрика за просторна структурна сличност сме утврдиле дека поднизата  $Y_k Y_{k+1} \dots Y_{k+n-2} Y_{k+n-1}$  е со најголем процент на просторна структурна сличност со протеинот X.

Аналогно на претходното разгледување се и разгледувањата кога должината на протеинот  $Y$  е помала од должината на протеинот  $X$  и кога нивните должини се еднакви. Претходниот пристап на просторно споредување може да се примени и за секоја подсеквенца од протеинот  $X$  со должина  $l, l \in 1, 2, \dots, n-1, n$ , за да се најде подсеквенца во рамки на протеинот  $Y$  со најголем процент на просторна структурна сличност.



## Стохастички модел за конкретна структура на протеин од иста протеинска фамилија

### 6.1 Предмет на анализа

Досегашната научна биоинформатичка анализа на структурата на протеините има утврдено дека протеините со слична примарна структура најчесто се хомологни протеини(имаат заеднички предходник) и припаѓаат на иста протеинска фамилија.

Сличната примарна структура на протеините од иста протеинска фамилија и фактот дека постојано се откриваат нови протеини кои припаѓаат кон одредена протеинска фамилија ми даде за идеја да предложам стохастички модел кој врз база на анализа на примарната структура на тековно откриените протеини во рамки на една протеинска фамилија ја определува веројатноста за конкретна примарна структура на протеин кој би припаѓал кон истата протеинска фамилија. Предложениот стохастички модел е конструиран врз база на разгледување на протеините како маркови системи, кои се карткомемориски системи чии наредни состојби зависат само од тековните состојби, но не и од претходните состојби.

## 6.2 Дискретни маркови вериги

Разгледуваме случаен процес  $\{X_n, n = 0, 1, 2, \dots\}$  кој е опишан со случајната променлива  $X$  која во дискретното време  $n$  ја дефинираме како  $X_n$  и истата може да прими било која вредност од множеството  $A = \{1, 2, \dots, R-1, R\}$  со одредена веројатност. За случајниот процес велиме дека во време  $n$  е во состојба  $i$ , каде  $i$  е елемент од множеството  $A$  ако и само ако случајната променлива  $X$  во време  $n$ , а тоа е  $X_n$ , прима вредност  $i$ , и пишуваме  $X_n = i$ .

Случајниот процес е дискретен марков процес (дискретна маркова верига) ако за истиот важи дека идната состојба на процесот е зависна само од тековната состојба, но не и од претходните состојби, односно важи:  $P(X_{n+1} = j \mid X_n = i, X_{n-1} = k_{n-1}, \dots, X_0 = k_0) = P(X_{n+1} = j \mid X_n = i) = P_{i,j}$ . Значи процесот во наредното дискретно време може да си ја задржи тековната состојба или да премине во друга состојба и таквиот премин или останување во тековната состојба е стохастички определен со условната веројатност  $P_{i,j}$ . Условните веројатности  $P_{i,j}$  се нарекуваат веројатности на едночекорен премин. Ако се познати сите веројатности на едночекорни премини може да се состави и матрица на веројатности на едночекорни премини. Матрицата го има обликот:

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,R-1} & P_{1,R} \\ P_{2,1} & P_{2,2} & \dots & P_{2,R-1} & P_{2,R} \\ \dots & \dots & \dots & \dots & \dots \\ P_{R,1} & P_{R,2} & \dots & P_{R,R-1} & P_{R,R} \end{bmatrix}. \quad \text{Елементот } P_{1,1} \text{ од матрицата } P \text{ ја претставува веројатноста ако}$$

процесот тековно се наоѓа во состојба 1 во наредното дискретно време повторно да се наоѓа во истата состојба, елементот  $P_{1,2}$  ја претставува веројатноста ако процесот тековно се наоѓа во состојба 1 во наредното дискретно време да се наоѓа во состојба 2, ... итн. Бидејќи марковиот процес од една состојба во наредното дискретно време може да премине во друга состојба или да остане во

истата, за истиот важи:  $\sum_{j=1}^R P_{i,j} = 1$ , за сите можни  $i$ .

## 6.2.1 Веројатности на n-чекорни премини и чапман-колмогорова равенка

Веројатноста дискретен марков процес после  $n$  премини ( $n$  дискретни времиња) од состојба  $i$  да премине во состојба  $j$  се нарекува веројатност на  $n$ -чекорен премин и истата се означува со  $P_{i,j}^{(n)}$ . Веројатностите на  $n$ -чекорни премини се пресметуваат со примена на чапман-колмогоровата равенка. Ако дискретиот марков процес може да се најде во  $R$  различни состојби тогаш веројатноста истиот после  $n$  премини од состојба  $i$  да премине во состојба  $j$  е:

$$P_{i,j}^{(n)} = P_{i,j}^{(s+m)} = \sum_{k=1}^R P_{i,k}^{(s)} P_{k,j}^{(m)}, \text{ за секое } s \text{ и } m \text{ така што } s \geq 1, s < n, m \geq 1, m < n \text{ и } s + m = n.$$

Со примена на чапман-колмогоровата равенка  $P_{i,j}^{(n)} = P_{i,j}^{(s+m)} = \sum_{k=1}^R P_{i,k}^{(s)} P_{k,j}^{(m)}$  можат да се пресметат

веројатностите на  $n$ -чекорни премини за сите можни парови на состојби  $(i, j)$  од кои понатаму се

составува матрица на веројатности на  $n$ -чекорни премини  $P^{(n)} = \begin{bmatrix} P_{1,1}^{(n)} & P_{1,2}^{(n)} & \dots & P_{1,R-1}^{(n)} & P_{1,R}^{(n)} \\ P_{2,1}^{(n)} & P_{2,2}^{(n)} & \dots & P_{2,R-1}^{(n)} & P_{2,R}^{(n)} \\ \dots & \dots & \dots & \dots & \dots \\ P_{R,1}^{(n)} & P_{R,2}^{(n)} & \dots & P_{R,R-1}^{(n)} & P_{R,R}^{(n)} \end{bmatrix}$ . Ако ги

имаме пресметано матрицата на веројатности на  $s$ -чекорни премини  $P^{(s)}$  и матрицата на веројатности на  $m$ -чекорни премини  $P^{(m)}$   $s \geq 1, s < n, m \geq 1, m < n$  и  $s + m = n$ , тогаш со примена на

чапман-колмогоровата равенка  $P_{i,j}^{(n)} = P_{i,j}^{(s+m)} = \sum_{k=1}^R P_{i,k}^{(s)} P_{k,j}^{(m)}$  за сите можни парови  $(i, j)$  од

множеството на можни состојби следи дека матрицата на веројатности на  $n$ -чекорни премини се добива со множење на матриците  $P^{(s)}$  и  $P^{(m)}$ , односно  $P^{(n)} = P^{(s+m)} = P^{(s)} \cdot P^{(m)}$ . Со примена на претходното матрицата на веројатности на  $n$ -чекорни премини може да се напише во облик:

$P^{(n)} = P \cdot P^{(n-1)} = P \cdot P \cdot P^{(n-2)} = \dots = P \cdot P \cdots P \cdot P\{n - \text{пати}\}$ , каде  $P$  е матрица на веројатности на едночекорни премини.

## 6.2.2 Пристапливи состојби и иредуцибилни дискретни маркови вериги

**Дефиниција:** За состојбата  $j$  велиме дека е пристаплива од состојбата  $i$  ако постои барем една веројатност поголема од нула процесот од состојба  $i$  после  $n$  премини да се најде во состојба  $j$ , односно барем за едно  $n > 0$  мора да важи  $P^{(n)}_{i,j} > 0$ .

**Дефиниција:** За состојбите  $i$  и  $j$  велиме дека комуницираат меѓусебно или дека образуваат комуницирачки пар ако состојбата  $j$  е пристаплива од состојбата  $i$  и состојбата  $i$  е пристаплива од состојбата  $j$ .

Ако состојбите  $i$  и  $j$  комуницираат меѓу себе - односно состојбата  $i$  е пристаплива од состојбата  $j$  и обратно тогаш таа релација меѓу истите се означува со  $\leftrightarrow$ , и пишуваме  $i \leftrightarrow j$ . Важи и следново: Ако состојбите  $i$  и  $j$  образуваат комуницирачки пар и состојбите  $j$  и  $k$  исто така образуваат комуницирачки пар, тогаш состојбите  $i$  и  $k$  исто така образуваат комуницирачки пар, односно ако  $i \leftrightarrow j$  и  $j \leftrightarrow k$ , тогаш  $i \leftrightarrow k$ .

**Дефиниција:** За дискретната маркова верига(процес) велиме дека е иредуцибилна дискретна маркова верига ако множеството на можни состојби за истата претставува комуницирачка класа - односно процесот од било која состојба може да се најде во било друга состојба во рамки на дозволеното множество.

### 6.2.3 Периодични и неперидични состојби на дискретна маркова верига

За состојбата  $i$  велиме дека е периодична со периода  $k (k > 1)$ , ако враќањето во таа состојба се случува после  $n \cdot k$  премини (дискретни времиња), при што  $n \in \mathbb{N}$ . Значи ако состојбата  $i$  е периодична со периода  $k$  тогаш дискретниот марков процес од состојба  $i$  по  $k$  или  $2k$  или  $3k, \dots$  итн премини повторно се враќа во состојбата  $i$ .

Важно што треба да забелижаме овде е тоа дека иако состојбата  $i$  може да биде периодична со периода  $k$  сепак можно е да не биде возможно процесот од состојбата  $i$  да се врати во истата за некои вредности на  $n \cdot k$ . Така на пример иако состојбата 1 да кажеме е периодична со периода 2 сепак може да биде невозможно да се врати во таа состојба после 2 или 6 премини но сепак враќањето во таа состојба мора да се случи по некои  $n \cdot 2$  премини за  $n \neq 1, n \neq 3$ .

Ако состојбите  $i$  и  $j$  се периодични и образуваат комуницирачки пар - односно  $i < - > j$ , тогаш периодата на враќање на процесот кон состојбата  $i$  која ја означуваме со  $d(i)$  е еднаква на периодата на враќање на процесот во состојбата  $j$  која ја означуваме со  $d(j)$ , односно  $d(i) = d(j)$ . Ако  $k = 1$  тогаш состојбата  $i$  е неперидична, односно не можеме да кажеме дека после одреден број на премини процесот повторно мора да се врати во состојбата  $i$ .

## 6.2.4 Рекурентни и транзиентни состојби кај дискретна маркова верига

За состојбата  $i$  велиме дека е транзиентна ако постои ненулта веројатност процесот откако ќе ја напушти никогаш повеќе да не се врати во истата. Односно ако со  $T_i$  ја означиме случајната променлива за временскиот период на повторно враќање на процесот во состојбата  $i$ , тогаш ако состојбата  $i$  е транзиентна постои веројатност поголема од нула процесот никогаш повеќе да не се врати во истата, односно веројатноста дека  $T_i = \infty$  е поголема од нула -  $P\{T_i = \infty\} > 0$ . Ако процесот после одреден број на премини откако ќе ја напушти состојбата  $i$  повторно се враќа во истата тогаш за истата состојба велиме дека е рекурентна - односно  $P\{T_i = \infty\} = 0$ .

Важно е да се каже дека постојат два типа на рекурентни состојби кај дискретните маркови вериги и тоа :

- Позитивно рекурентни
- Нулто рекурентни

За да ја објасниме разликата помеѓу истите потребно е да се дефинира очекувано време на враќање на процесот кон рекурентната состојба кое се бележи со  $M_i = E[T_i]$ . Ако очекуваното време на враќање на процесот во состојбата  $i$  -  $M_i = E[T_i]$  е конечно тогаш велиме дека состојбата  $i$  е позитивно рекурентна, во спротивност истата е нулто рекурентна.

Исто така освен тоа што дадена состојба  $i$  може да биде транзиентна или рекурентна исто така таа може да биде и апсорбирачка, при што доколку процесот се најде во состојба која сме ја дефинирале како апсорбирачка тогаш истиот никако не може да ја напушти истата - односно важи:

$$P_{i,i} = 1, P_{i,j} = 0, i \neq j.$$

### 6.3 Опис на моделот

Протеините се секвенци изградени од дваесет различни аминокиселини. Листата на аминокиселини кои влегуваат во градбата на протеините и нивните симболи се дадени во Табела 6.3.1. Примарното ниво на организација на протеините ни овозможува истите да ги разгледуваме како низи од аминокиселини, каде секоја позиција на таа низа е пополнета со една од дваесетте аминокиселини дадени во Табела 6.3.1. Ако со  $AC=\{A,R,N,D,C,E,Q,G,P,S,I,H,L,M,K,F,W,Y,T,V\}$  го означиме множеството на дваесетте аминокиселини кои влегуваат во градбата на протеините, тогаш секој откриен и верифициран протеин разгледуван на примарно ниво е низа од  $n$  аминокиселини  $\{a_i\}_{i=1}^n = a_1, a_2, \dots, a_{n-1}, a_n$ , каде секоја од аминокиселините  $a_i$  е елемент од множеството  $AC$ .

Аминокиселина	Симбол
Alanine	A
Arginine	R
Asparagine	N
Aspartic acid	D
Cysteine	C
Glutamic acid	E
Glutamine	Q
Glycine	G
Proline	P
Serine	S
Isoleucine	I
Histidine	H
Leucine	L
Methionine	M
Lysine	K
Phenylalanine	F
Tryptophan	W
Tyrosine	Y
Threonine	T
Valine	V

Табела 6.3.1 Листа на аминокиселини

Разгледуваме протеинска фамилија која во општост се состои од  $k$  протеини,  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$ . Секој од протеините  $\Pi_i$  содржи позитивен целоброен број на аминокиселини, што значи дека секој протеин  $\Pi_i$  е низа од  $n$  аминокиселини  $\Pi_i = \{a_i\}_{i=1}^n$ , дефинирана над множеството од дваесет градобени аминокиселини **АС**.

Стохастичкото разгледување на протеините во рамки на кое веројатноста на позиција  $j+1$  да се наоѓа конкретна аминокиселина  $a_{j+1}$  зависи само од аминокиселината  $a_j$  која се наоѓа на позиција  $j$ , но не и од аминокиселините  $a_{j-1}, a_{j-2}, \dots, a_2, a_1$  на позиции  $j-1, j-2, \dots, 2, 1$ , ни овозможува протеините на примарно ниво да можеме ги разгледуваме како маркови системи. Во општост марков системи е секој караткомемориски систем кај кој наредната состојба на системот зависи само од тековната, но не и од претходните состојби во кои се наоѓал системот.

Ако со  $a_{j+1}$  ја означиме аминокиселината на позиција  $j+1$ , со  $a_j$  ја означиме аминокиселината на позиција  $j$  и со  $a_{j-1}, a_{j-2}, \dots, a_2, a_1$  ги означиме аминокиселините на позиции  $j-1, j-2, \dots, 2, 1$  кај протеинот  $\Pi_i$ , кога истиот се разгледува како марков систем за истиот важи:

$$P(a_{j+1} | a_j, a_{j-1}, \dots, a_2, a_1) = P(a_{j+1} | a_j).$$

За секој протеин  $\Pi_i$  од протеинската фамилија  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$  можеме да конструираме фрекфентна квадратна матрица  $\mathbf{F}_i$  со дваесет колони и дваесет редици, кај која секој елемент  $f_{X,Y} = \frac{n_{X,Y}}{n_X}$ , каде  $n_{X,Y}$  е број на појавувања на аминокиселината  $Y$  како наредна по аминокиселината  $X$  во рамки на протеинот  $\Pi_i$  а  $n_X$  е вкупниот број на појавувања на аминокиселината  $X$  во рамки на протеинот  $\Pi_i$  како нетерминална аминокиселина (по истата следи наредна аминокиселина од протеинот  $\Pi_i$ ). Структурата на фрекфентната квадратна матрица  $\mathbf{F}$  е дадена во продолжение.



	A	R	N	D	C	E	Q	G	P	S	I	H	L	M	K	F	W	Y	T	V
A	$f_{A,A}$	$f_{A,R}$	$f_{A,N}$	$f_{A,D}$	$f_{A,C}$	$f_{A,E}$	$f_{A,Q}$	$f_{A,G}$	$f_{A,P}$	$f_{A,S}$	$f_{A,I}$	$f_{A,H}$	$f_{A,L}$	$f_{A,M}$	$f_{A,K}$	$f_{A,F}$	$f_{A,W}$	$f_{A,Y}$	$f_{A,T}$	$f_{A,V}$
R	$f_{R,A}$	$f_{R,R}$	$f_{R,N}$	$f_{R,D}$	$f_{R,C}$	$f_{R,E}$	$f_{R,Q}$	$f_{R,G}$	$f_{R,P}$	$f_{R,S}$	$f_{R,I}$	$f_{R,H}$	$f_{R,L}$	$f_{R,M}$	$f_{R,K}$	$f_{R,F}$	$f_{R,W}$	$f_{R,Y}$	$f_{R,T}$	$f_{R,V}$
N	$f_{N,A}$	$f_{N,R}$	$f_{N,N}$	$f_{N,D}$	$f_{N,C}$	$f_{N,E}$	$f_{N,Q}$	$f_{N,G}$	$f_{N,P}$	$f_{N,S}$	$f_{N,I}$	$f_{N,H}$	$f_{N,L}$	$f_{N,M}$	$f_{N,K}$	$f_{N,F}$	$f_{N,W}$	$f_{N,Y}$	$f_{N,T}$	$f_{N,V}$
D	$f_{D,A}$	$f_{D,R}$	$f_{D,N}$	$f_{D,D}$	$f_{D,C}$	$f_{D,E}$	$f_{D,Q}$	$f_{D,G}$	$f_{D,P}$	$f_{D,S}$	$f_{D,I}$	$f_{D,H}$	$f_{D,L}$	$f_{D,M}$	$f_{D,K}$	$f_{D,F}$	$f_{D,W}$	$f_{D,Y}$	$f_{D,T}$	$f_{D,V}$
C	$f_{C,A}$	$f_{C,R}$	$f_{C,N}$	$f_{C,D}$	$f_{C,C}$	$f_{C,E}$	$f_{C,Q}$	$f_{C,G}$	$f_{C,P}$	$f_{C,S}$	$f_{C,I}$	$f_{C,H}$	$f_{C,L}$	$f_{C,M}$	$f_{C,K}$	$f_{C,F}$	$f_{C,W}$	$f_{C,Y}$	$f_{C,T}$	$f_{C,V}$
E	$f_{E,A}$	$f_{E,R}$	$f_{E,N}$	$f_{E,D}$	$f_{E,C}$	$f_{E,E}$	$f_{E,Q}$	$f_{E,G}$	$f_{E,P}$	$f_{E,S}$	$f_{E,I}$	$f_{E,H}$	$f_{E,L}$	$f_{E,M}$	$f_{E,K}$	$f_{E,F}$	$f_{E,W}$	$f_{E,Y}$	$f_{E,T}$	$f_{E,V}$
Q	$f_{Q,A}$	$f_{Q,R}$	$f_{Q,N}$	$f_{Q,D}$	$f_{Q,C}$	$f_{Q,E}$	$f_{Q,Q}$	$f_{Q,G}$	$f_{Q,P}$	$f_{Q,S}$	$f_{Q,I}$	$f_{Q,H}$	$f_{Q,L}$	$f_{Q,M}$	$f_{Q,K}$	$f_{Q,F}$	$f_{Q,W}$	$f_{Q,Y}$	$f_{Q,T}$	$f_{Q,V}$
G	$f_{G,A}$	$f_{G,R}$	$f_{G,N}$	$f_{G,D}$	$f_{G,C}$	$f_{G,E}$	$f_{G,Q}$	$f_{G,G}$	$f_{G,P}$	$f_{G,S}$	$f_{G,I}$	$f_{G,H}$	$f_{G,L}$	$f_{G,M}$	$f_{G,K}$	$f_{G,F}$	$f_{G,W}$	$f_{G,Y}$	$f_{G,T}$	$f_{G,V}$
P	$f_{P,A}$	$f_{P,R}$	$f_{P,N}$	$f_{P,D}$	$f_{P,C}$	$f_{P,E}$	$f_{P,Q}$	$f_{P,G}$	$f_{P,P}$	$f_{P,S}$	$f_{P,I}$	$f_{P,H}$	$f_{P,L}$	$f_{P,M}$	$f_{P,K}$	$f_{P,F}$	$f_{P,W}$	$f_{P,Y}$	$f_{P,T}$	$f_{P,V}$
S	$f_{S,A}$	$f_{S,R}$	$f_{S,N}$	$f_{S,D}$	$f_{S,C}$	$f_{S,E}$	$f_{S,Q}$	$f_{S,G}$	$f_{S,P}$	$f_{S,S}$	$f_{S,I}$	$f_{S,H}$	$f_{S,L}$	$f_{S,M}$	$f_{S,K}$	$f_{S,F}$	$f_{S,W}$	$f_{S,Y}$	$f_{S,T}$	$f_{S,V}$
I	$f_{I,A}$	$f_{I,R}$	$f_{I,N}$	$f_{I,D}$	$f_{I,C}$	$f_{I,E}$	$f_{I,Q}$	$f_{I,G}$	$f_{I,P}$	$f_{I,S}$	$f_{I,I}$	$f_{I,H}$	$f_{I,L}$	$f_{I,M}$	$f_{I,K}$	$f_{I,F}$	$f_{I,W}$	$f_{I,Y}$	$f_{I,T}$	$f_{I,V}$
H	$f_{H,A}$	$f_{H,R}$	$f_{H,N}$	$f_{H,D}$	$f_{H,C}$	$f_{H,E}$	$f_{H,Q}$	$f_{H,G}$	$f_{H,P}$	$f_{H,S}$	$f_{H,I}$	$f_{H,H}$	$f_{H,L}$	$f_{H,M}$	$f_{H,K}$	$f_{H,F}$	$f_{H,W}$	$f_{H,Y}$	$f_{H,T}$	$f_{H,V}$
L	$f_{L,A}$	$f_{L,R}$	$f_{L,N}$	$f_{L,D}$	$f_{L,C}$	$f_{L,E}$	$f_{L,Q}$	$f_{L,G}$	$f_{L,P}$	$f_{L,S}$	$f_{L,I}$	$f_{L,H}$	$f_{L,L}$	$f_{L,M}$	$f_{L,K}$	$f_{L,F}$	$f_{L,W}$	$f_{L,Y}$	$f_{L,T}$	$f_{L,V}$
M	$f_{M,A}$	$f_{M,R}$	$f_{M,N}$	$f_{M,D}$	$f_{M,C}$	$f_{M,E}$	$f_{M,Q}$	$f_{M,G}$	$f_{M,P}$	$f_{M,S}$	$f_{M,I}$	$f_{M,H}$	$f_{M,L}$	$f_{M,M}$	$f_{M,K}$	$f_{M,F}$	$f_{M,W}$	$f_{M,Y}$	$f_{M,T}$	$f_{M,V}$
K	$f_{K,A}$	$f_{K,R}$	$f_{K,N}$	$f_{K,D}$	$f_{K,C}$	$f_{K,E}$	$f_{K,Q}$	$f_{K,G}$	$f_{K,P}$	$f_{K,S}$	$f_{K,I}$	$f_{K,H}$	$f_{K,L}$	$f_{K,M}$	$f_{K,K}$	$f_{K,F}$	$f_{K,W}$	$f_{K,Y}$	$f_{K,T}$	$f_{K,V}$
F	$f_{F,A}$	$f_{F,R}$	$f_{F,N}$	$f_{F,D}$	$f_{F,C}$	$f_{F,E}$	$f_{F,Q}$	$f_{F,G}$	$f_{F,P}$	$f_{F,S}$	$f_{F,I}$	$f_{F,H}$	$f_{F,L}$	$f_{F,M}$	$f_{F,K}$	$f_{F,F}$	$f_{F,W}$	$f_{F,Y}$	$f_{F,T}$	$f_{F,V}$
W	$f_{W,A}$	$f_{W,R}$	$f_{W,N}$	$f_{W,D}$	$f_{W,C}$	$f_{W,E}$	$f_{W,Q}$	$f_{W,G}$	$f_{W,P}$	$f_{W,S}$	$f_{W,I}$	$f_{W,H}$	$f_{W,L}$	$f_{W,M}$	$f_{W,K}$	$f_{W,F}$	$f_{W,W}$	$f_{W,Y}$	$f_{W,T}$	$f_{W,V}$
Y	$f_{Y,A}$	$f_{Y,R}$	$f_{Y,N}$	$f_{Y,D}$	$f_{Y,C}$	$f_{Y,E}$	$f_{Y,Q}$	$f_{Y,G}$	$f_{Y,P}$	$f_{Y,S}$	$f_{Y,I}$	$f_{Y,H}$	$f_{Y,L}$	$f_{Y,M}$	$f_{Y,K}$	$f_{Y,F}$	$f_{Y,W}$	$f_{Y,Y}$	$f_{Y,T}$	$f_{Y,V}$
T	$f_{T,A}$	$f_{T,R}$	$f_{T,N}$	$f_{T,D}$	$f_{T,C}$	$f_{T,E}$	$f_{T,Q}$	$f_{T,G}$	$f_{T,P}$	$f_{T,S}$	$f_{T,I}$	$f_{T,H}$	$f_{T,L}$	$f_{T,M}$	$f_{T,K}$	$f_{T,F}$	$f_{T,W}$	$f_{T,Y}$	$f_{T,T}$	$f_{T,V}$
V	$f_{V,A}$	$f_{V,R}$	$f_{V,N}$	$f_{V,D}$	$f_{V,C}$	$f_{V,E}$	$f_{V,Q}$	$f_{V,G}$	$f_{V,P}$	$f_{V,S}$	$f_{V,I}$	$f_{V,H}$	$f_{V,L}$	$f_{V,M}$	$f_{V,K}$	$f_{V,F}$	$f_{V,W}$	$f_{V,Y}$	$f_{V,T}$	$f_{V,V}$

Слика 6.2.2 Фрекфентна квадратна матрица F

За низата на протеини  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$  од протеинската фамилија може да се конструира низа од фрекфентни квадратни матрици  $F_1, F_2, \dots, F_{k-1}, F_k$ , каде за секој протеин  $\Pi_i$  е конструирана фрекфентна квадратна матрица  $F_i$ . Со собирање на фрекфентните квадратни матрици  $F_1, F_2, \dots, F_{k-1}, F_k$  и делење на секој од елементите на резултантната матрица со бројот на протеини во рамки на протеинската фамилија сме ја определиле матрицата  $P = \frac{1}{k} \sum_{i=1}^k F_i$ , каде секој елемент  $P_{X,Y}$  во рамки на матрицата  $P$  ја претставува просечната веројатност како наредна аминокиселина по аминокиселината  $X$  да следи аминокиселината  $Y$  кај протеин кој би припаѓал кон тековната протеинска фамилија  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$ .

За конструкција на стохастичкиот модел потребно е и да се определи веројатноста протеинот да започнува со конкретна аминокиселина од множеството на градбени аминокиселини  $AC$ . Таа веројатност може да се определи на два начини дадени во продолжение.

Ако веројатноста за појавување на конкретна аминокиселина како почетна ја утврдуваме само врз база на анализа на почетните аминокиселини кај протеините од протеинската фамилија, тогаш според овој пристап веројатноста конкретна аминокиселина  $X$  да биде почетна за протеин од истата

протеинска фамилија е  $p_X = \frac{n_X}{k}$ , каде  $n_X$  е бројот на појавувања на аминокиселината  $X$  како

почетна во рамки на протеинската фамилија  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$ . Другиот пристап на утврдување на веројатноста за појавување на конкретна аминокиселина како почетна се базира на анализа на сите аминокиселини во рамки на протеинската фамилија. За секоја аминокиселина  $X$  од множеството на

градбени аминокиселини  $AC$  можат да се определат вредностите:  $\frac{n_{X,1}}{l(\Pi_1)}, \frac{n_{X,2}}{l(\Pi_2)}, \dots, \frac{n_{X,k-1}}{l(\Pi_{k-1})}, \frac{n_{X,k}}{l(\Pi_k)}$ ,

каде секој од елементите  $n_{X,i}$  го претставува бројот на појавувања на аминокиселината  $X$  кај протеинот  $\Pi_i$  а  $l(\Pi_i)$  е бројот на аминокиселини кои влегуваат во градбата на протеинот  $\Pi_i$ . Според овој пристап веројатноста конкретна аминокиселина  $X$  да се појави како почетна е:

$$p_X = \frac{\frac{n_{X,1}}{l(\Pi_1)} + \frac{n_{X,2}}{l(\Pi_2)} + \dots + \frac{n_{X,k-1}}{l(\Pi_{k-1})} + \frac{n_{X,k}}{l(\Pi_k)}}{k}.$$

По наоѓање на матрицата  $\mathbf{P}$  и веројатностите  $p_X$  за секоја од двесетте протеински градбени аминокиселини, можеме да ја дадеме генералната формула за веројатност на конкретна примарна структура на протеин од иста протеинска фамилија, базирана на разгледување на протеините како маркови системи.

Веројатноста за конкретна примарна структура  $X_1, X_2, \dots, X_{n-1}, X_n$  на протеин кој би припаѓал кон протеинската фамилија  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$  од која се изведени матрицата  $\mathbf{P}$  и веројатностите  $p_X$  за појавување на секоја од аминокиселините како почетни во рамки на протеин кој би припаѓал кон таа протеинска фамилија е:

$$P(X_1, X_2, \dots, X_{n-1}, X_n) = P(X_n | X_{n-1}, X_{n-2}, \dots, X_2, X_1) \cdot P(X_{n-1} | X_{n-2}, X_{n-3}, \dots, X_2, X_1) \cdots P(X_3 | X_2, X_1) \cdot P(X_2 | X_1) \cdot P(X_1).$$

Со примена на концептот на разгледување на протеините на примарно ниво како маркови системи секоја од веројатностите  $P(X_i | X_{i-1}, X_{i-2}, \dots, X_2, X_1)$  зависи само од елементите  $X_{i-1}$ , но не и од структурата на низата  $X_{i-2}, X_{i-3}, \dots, X_2, X_1$ , што во конкретниот случај би значело дека веројатноста на позиција  $i$  кај протеинот да се јави конкретна аминокиселина  $X_i$  од множеството на дваесет градбени аминокиселини е зависна само од аминокиселината  $X_{i-1}$  на позиција  $i-1$  но не и од претходната структура на протеинот (структурата на поднизата  $X_{i-2}, X_{i-3}, \dots, X_2, X_1$ ), ни овозможува секоја од веројатностите  $P(X_i | X_{i-1}, X_{i-2}, \dots, X_2, X_1)$  да можат да бидат напишани во облик:

$$P(X_i | X_{i-1}, X_{i-2}, \dots, X_2, X_1) = P(X_i | X_{i-1}).$$

Со ова равенството за веројатност на конкретна примарна структура  $X_1, X_2, \dots, X_{n-1}, X_n$  на протеин од иста протеинска фамилија го добива обликот:

$$P(X_1, X_2, \dots, X_{n-1}, X_n) = P(X_n | X_{n-1}) \cdot P(X_{n-1} | X_{n-2}) \cdots P(X_3 | X_2) \cdot P(X_2 | X_1) \cdot P(X_1).$$

Веројатностите на позиција  $i$  да се наоѓа конкретна аминокиселина  $X_i$  ако на позиција  $i-1$  се наоѓа аминокиселината  $X_{i-1}$  кај протеин кој би припаѓал кон тековната протеинска фамилија  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$  се еднакви на просечните веројатности како наредна аминокиселина по аминокиселината  $X_{i-1}$  да следува аминокиселината  $X_i$ , утврдени врз база на анализа на примарната структура на тековно откриените и верифицирани протеини  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$ , односно важи  $P(X_i | X_{i-1}) = P_{X_{i-1}, X_i}$ ,  $i \in 2, 3, \dots, n-1, n$ . Веројатноста аминокиселината  $X_1$  да се јави како почетна  $P_{X_1}$  може да се определи на начините предложени претходно, од каде следи дека веројатноста за конкретната примарна структура  $X_1, X_2, \dots, X_{n-1}, X_n$  на протеин кој би припаѓал кон протеинската фамилија  $\Pi_1, \Pi_2, \dots, \Pi_{k-1}, \Pi_k$  е  $P(X_1, X_2, \dots, X_{n-1}, X_n) = P_{X_{n-1}, X_n} \cdot P_{X_{n-2}, X_{n-1}} \cdot P_{X_{n-3}, X_{n-2}} \cdots P_{X_2, X_3} \cdot P_{X_1, X_2} \cdot P_{X_1}$ .

Претходното равенство е генералната формула која го опишува моделот за веројатност на конкретна примарна структура на протеин кој би припаѓал кон конкретна протеинска фамилија конструиран врз база на анализа на примарната структура на сите тековно откриени и верифицирани протеини од таа протеинска фамилија.

# Заклучок

Овој магистерски труд е производ на неколкумесечни истражувања и анализи на структурата на протеините на различни нивоа на организација. Покрај деталниот осврт за структурата на протеините и најупотребуваните модели и методи за претстава на структурата на протеините, предложени се и два нови модела. Првиот модел е векторски базиран модел и истиот ја опишува просторната структура на скелетот на протеинот, додека пак вториот модел е стохастички базиран модел кој ја предвидува конкретната примарна структура на протеин кој би припаѓал кон одредена протеинска фамилија.

Скелетниот модел на множеството на скелетни атоми со хаотично распределени координатни позиции му дава математичка просторно-логичка димензија со примена на напреднати концепти од векторска анализа. Како модел се карактеризира со низа на параметри, при што неколкумесечната примена на овој модел на голем број на протеини од различни протеински фамилии утврди константност на дел од параметрите и променливост на останатите. Таквата анализа беше потпомогната со самостојно напишан програмски C++ код за автоматизација на низата на алгебарски, векторски и статистички пресметки.

Во рамки на овој магистерски труд беше предложен и стохастички модел за веројатност на примарна структура на протеин кој би припаѓал кон одредена протеинска фамилија. Таквиот модел е изведен врз база на разгледување на низата на тековно откриени и верифицирани протеини од една протеинска фамилија како маркови системи, при што генералните резултати од тоа разгледување за протеинската фамилија се употребени за формирање на матрица на веројатности за промена на состојба од која се изведува генералната формула за веројатност на конкретна примарна структура на протеин кој би припаѓал кон таа протеинска фамилија.

Сметам дека предложените модели со применетата софтверска анализа и генералните заклучоци во однос на резултатите од истата се своевиден допринос кон биоинформатичката анализа на протеините и своевиден патокз за идна научно-истражувачка работа.

## Листа на користени трудови во истражувањето

- [1] Richard Durbin, Sean R. Eddy, Anders Krogh, Graeme Mitchison “Biological sequence analysis”, Cambridge University Press 1998
- [2] Soumya Raychaudhuri , “Computational Text Analysis for Functional Genomics and Bioinformatics”, Oxford University Press 2006
- [3] Hui-Huang Hsu, “Advanced Data mining Technologies in Bioinformatics” Tamkang Univerity, Taipei, Taiwan 2006
- [4] Peter E.Nielsen, “Nucleic Acids Methods and Protocol”, vol 208
- [5] Andrzej Polanski, Maerk Kimmerl, “Bioinformatics” , Springer-Verlag Berlin Heidelberg 2007
- [6] Essential Bioinformatics, “Jin Xiong” , Texas A&M University, Cambridge University Press 2006
- [7] Warren Ewens, Gregory Grant, “Statistical Methods in Bioinformatics”, Springer 2005
- [8] Jonatan Pevsner, “Bioinformatics and Functional Genomics”, John Wiley and Sons, Inc Publication
- [9] Ashburner, M., C.A.Ball et al.(2000). “Gene Ontology”
- [10] Blashke, C.,M.A.Andrade, et al.(1999), “Automatic extraction of biological information from scientific text: protein-protein interactions, Proc. Int. Conf. Intell Syst Mol.Biol.2(1): 60-7
- [11] Branden, C., and Tozze, J. 1999. “Introduction to Protein Structure”, 2<sup>nd</sup> edition. New York: Garland Publishing
- [12] Scheeff, E.D., and Fink, J. L. 2003. “Fundamentals of protein structure” In Structural Bioinformatics, edited by P. E. Bourne and H.Weissing, 15-39. Hoboken, NJ Wiley-Liss
- [13] Bourne, P.E., and Shindyalov, I.N., 2003 “Structure comparison and alignment”
- [14] Carugo, O., and Pongor, S., 2002. “Recent progress in Protein 3D structure comparison”
- [15] Kinch, L.N., and Grishin, N.V.2002. “Evolution of protein structures and functions”
- [16] Koehl, P.2001. “Protein structure similarities”
- [17] Reddy, B.J. 2003 “Protein structure evolution and scope database”

- [18] Russell, R.B. 2002. "Classification of protein folds" *Mol. Biotechnol.* 20: 17-28
- [19] Tate, J. 2003. "Molecular visualization"
- [20] Edwards, Y.J., and Cottage, A. 2003. "Bioinformatics methods to predict protein structure and functions" A practical approach. *Mol. Biotechnol.* 23: 139-66
- [21] Hering, J. 2002. "Computational methods for protein secondary structure prediction using multiple sequence alignments" *Curr. Protein Pept. Sci.* 1:273-301
- [22] Al-Lizikani, B., Jung, J., Xiang, Z., and Honig, B. 2001. "Protein structure prediction" *Curr. Opin. Chem. Biol.* 5:51-6
- [23] Baker, D., and Saly, A., 2001. "Protein structure prediction and structural genomics" *Science.* 294: 93-6
- [24] Bonneau, R., and Baker, D. 2001. "Ab initio protein structure prediction: Progress and prospects" *Annu. Rev. Biophys. Biomol. Struct.* 30: 173-89
- [25] Chivian, D., Robertson, T., Bonneau, R., and Baker, D. 2003. "Ab initio methods"
- [26] Fetrow, J.S., Giammona, A., Kolinski, A., and Skolnick, J. 2002., "The protein folding problem: A biophysical enigma" *Curr. Pharm. Biotechnol.* 3: 329-47
- [27] Forster, M.J. 2000. "Molecular modelling in structural biology" *Micron.* 33: 365-84
- [28] Godzik, A. 2003. "Fold recognition methods"
- [29] Krieger, E., Nabuurs, S.B., and Vriend, G. 2003. "Homology modeling"
- [30] Xu, D., Xu, Y., and Uberbacher, E.C. 200. "Computational tools for protein modeling" *Curr. Protein Pept. Sci.* 1: 1-21
- [31] Schulz, G. E., Schirmer, R.H., (1990), "Principles of Protein Structure", Springer
- [32] Sharke, A., Rupley, J.A., (1973), "Environment and exposure to solvent of protein atoms", *J. Mol. Biol.* vol 79
- [33] Simons, K. T., Bonneau, R., Ruczinski I., Baker, D., (1999) "Ab initio protein structure prediction of CASP3 target using ROSETTA"
- [34] Sander, C., Schneider, R., (1991) "Database of homology derived protein structures and the structural meaning of sequence alignment"
- [35] McLachan, A. D. (1982), "Rapid comparison of protein structures"
- [36] Iosifescu, M. (1990), "Finite Markov Processes and their Applications", Wiley

- [37] Baldi, P., Brunka, S., (1998) “Bioinformatics”, MIT press
- [38] Bowie, J. U., Luthi, R., Eisenberg, D., (1991) “A method to protein sequence that fold into a known three dimensional structure”, Science, vol. 253
- [39] Crick, F., (1970), “Central dogma of molecular biology”, Nature, vol. 227
- [40] Koski, T., Koskinen, T., (2001), “Hidden markov models for bioinformatics”, Kluwer Academic